Classification of COROT Exoplanet Light Curves

Progress report since CW9:
visit of L. Sarro to Leuven in March, including meeting of whole team in Leuven that week + common work ever since

L.M. Sarro
J. Debosscher
C. Aerts
R. Garrido
E. Solano
B. Vandenbussche

Outline

- Training classes
- Light curve analysis
- Preprocessing
- Classifiers
- Results
- To be done
- Conclusions

Training set (Supervised Classification)

- About 40 stellar variability classes (see CoRoT book)
- Extensive literature search for well known member star identifiers + consultation of CoRoT community
- Search for their light curves (public+private):
    - HIPPARCOS
    - Geneva
    - Strömgren
    - ESO LTPV
    - ULTRACAM
    - MOST
- Additional colour information: 2MASS

Light Curve Analysis

1) Check for trends (can also be indicative for long-term periodicity)
2) Implemented period analysis methods:
   - Fourier techniques (FFT and fast Lomb-Scargle)
     - Suited for pulsating variables
     - Search for up to a maximum of *3 significant different frequencies:* iterative procedure of peak selection + harmonic fit subtraction (4 harmonics)
     - *Mono-Multi periodicity flag* based on variance ratios (original variance/variance of residues after subtraction of harmonic fit with *f1)*
   - PDM better for highly non-sinusoidal light curves
     - More suitable for eclipsing binaries
     - Search for main frequency only

Preprocessing

- *Attribute selection: which attributes provide the best overall separation between classes? Which attributes will degrade the classifier's performance?*
- Statistical methods for attribute evaluation:
  - For single attributes
    - Information Gain
    - Principal Components
    - .....
  - For attribute subsets
    - Correlation based Feature Selection (CFS)
    - Wrapper approach

Classifiers

- Two levels:
  - *COROT classifier:* using only information delivered by COROT (light curve parameters)
  - *others classifier:* using additional attributes from other sources, e.g. 2 MASS colours, GAUDI,...

Classifiers (cont.)

- 5 Statistical Learning Methods are compared for the COROT classifier:
  1) Bayesian networks
  2) Bayesian ensembles of neural networks
  3) Support Vector Machines
  4) Decision Trees
  5) Rule Sets

- Performance tested with a provisional training set consisting of HIPPARCOS light curves with and without 2MASS colour information

First results (provisional training set)

- Attribute selection:
  - *Relevant*: frequencies (*f 1, f 2, f 3*), 2MASS colours, amplitudes, amplitude ratios between harmonics of *f 1*
  - *Irrelevant*: phase-differences, amplitude ratios between harmonics of *f 2* and *f 3* (with a few exceptions)
- Classifier performance (statistical):
  - Best: Bayesian ensembles of Neural Networks
  - Equivalent: Bayesian Networks, Decision Trees and Rule sets
  - Worst: Support Vector Machines
  - Average best performance increases from 65 to 75% correct classifications when additional colour information is used (2MASS)

Future work

- Update training set for the 40 classes (~June).
- Construct more specific classifiers for different regions in HR-diagram
- Identify problematic regions in attribute space (source of misclassification) + new attributes to remove possible degeneracies.
- Assess new measures of classifier performance
- Repeat the procedures for the updated training set
- Finalise routines in order to obtain results in 2-weeks timescale for CoRoT long run data

Conclusions

- Visit of J. Debosscher to Madrid planned in July 2006 to finalise set-up for N3 dataproduct
- We can perform preliminary classification and N3 flag at any stage upon request of CoRoT SC
- We are ready and awaiting the CoRoT data !!