Robust Machine Learning Applied to Terascale Astronomical Datasets: Summary

Nicholas M. Ball, Cosmic Cartography Poster II-D, Dec 6th 2007

ASTRONOMY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
NCSA
Laboratory for Cosmological Data Mining (LCDM)

The LCDM is a UIUC-NCSA collaboration to apply machine learning to large astronomical datasets and study cosmological statistics. We present:

- Classifications of the form P(galaxy, star, neither) for 143×10^6 SDSS objects using decision trees
- Improved quasar photometric redshifts without regions of 'catastrophic failure' using kNN
- Probability density functions in redshift for galaxies and quasars in the SDSS and SDSS+GALEX
- LRAC access to NCSA supercomputing allows powerful implementation
- All classifications and photometric redshifts are blind-tested
- PDFs allow much-improved subsamples of quasars
- Future work will be +IR and COSMOS, possibly FPGAs and GPUs
- For more information see Ball et al. 2006, 2007a, b (^Ball, N on ADS or http://nball.astro.uiuc.edu under publications; 2007b is in prep)