Poster at Splinter Meeting

Splinter G

Data mining in the spectra database of the Sloan Digital Sky Survey by means of Kohonen self-organising maps

H. Meusinger[1,2], P. Schalldach[1], N. Balafkan[3], F. Pertermann[1], J. Brünecke[2], A. in der Au[4], J. Schwarzer[5]

[1] *Thüringer Landessternwarte Tautenburg*
[2] *University Leipzig, Faculty of Physics and Geosciences*
[3] *Saint Mary's University, Halifax, Canada, Department of Astronomy and Physics*
[4] *Texture Editor, Munich*
[5] *Systemberatung Schwarzer, Bonn*

Over the past 15 years, the Sloan Digital Sky Survey (SDSS) I-III has obtained four million spectra, mostly from galaxies and quasars. Mining such a tremendous data pool must lead to the discovery of very rare spectral types. However, whereas the spectroscopic pipeline of the SDSS is accurate and efficient for the vast majority, it fails in case of unusual spectra. We developed the software package ASPECT that is able to organise large spectra data pools by means of similarity in a topological map. The approach is based on the Kohonen method of self-organising maps (SOMs), an artificial neural network algorithm that uses unsupervised learning to produce a two-dimensional mapping of higher-order input data. The resulting SOM with its clustering properties constitutes an efficient tool for the selection of certain spectral types providing simultaneously a greater picture of the entire data set. We computed a huge SOM for one million spectra from the SDSS Data Release (DR) 7 and hundreds of smaller SOMs for galaxy and quasar spectra from SDSS DR10 binned in narrow redshift intervals. The computation of a SOM for all spectra from the final SDSS III data release, which is more demanding in terms of hardware and software, is in preparation. So far the SOMs were applied mainly for the search of rare spectral types: odd quasar spectra such as weak line quasars or unusual broad absorption line quasars (details are given in the talk in Splinter A), E+A galaxies, supernovae, and C stars. Other applications of SOMs computed with ASPECT are also possible: clustering of photometric spectral energy distributions or of structure functions from light curves.