

Complexity Science DTC Miniproject

Automated classification of astronomical spectra using artificial neural networks

B.T. Gänsicke, T.R. Marsh, D. Steeghs (Astronomy & Astrophysics)

With the advances in digital detector technologies, astronomical surveys are continuously growing in size and scope, the most prominent example being the Sloan Digital Sky Survey (SDSS, [1,2]), providing currently morphology and broad-band colours for nearly 300 million objects in the sky, as well as almost 1.3 million spectra (Figure 1). Vast data sets such as SDSS hold enormous potential for creating large well-defined samples of various types of astronomical objects that can be used for population studies, as well as for finding extremely rare objects that can unravel questions about the evolution of planets, stars, and galaxies. However, because of their sheer size, the ongoing large-scale surveys require the development of robust automated classification algorithms to fully exploit their scientific potential.

In previous work on automated spectral classification [e.g. 3,4] artificial neural networks (ANN), in particular the multilayer back propagation network [5], have been explored. ANN require an initial supervised training phase using well-defined template spectra in order to modify the weights among the neural connections between nodes in the different layers. Automated classification using ANN has been extended to pre-process the spectra by means of a principle component analysis (PCA), which compresses the spectral data and reduces the noise level while retaining the inherent information of the data [e.g. 6,7]. While the success rate of ANN-based classification appears to be sufficiently high, applications have so far been limited to data sets that were small in the number of spectra (a few 1000) and in breadth (in most occasions only consisting of normal stars, with the exception of [8]).

The aim of this MiniProject is to develop an algorithm that can automatically classify spectra from the SDSS data base. A specific goal will be the identification of an expected ~ 12000 white dwarfs in the SDSS spectral data base, that will be subjected to detailed analysis by the Warwick astronomy group. Besides acquiring knowledge on the theory and application of ANN, this MiniProject will provide insight into a number of astrophysical problems, as well as into methodologies employed in astronomy. This project can naturally lead onto a PhD focused on the design of more sophisticated automated classification algorithms that can e.g. process the entire $\simeq 1.3 \times 10^6$ spectra from SDSS, or $\simeq 15000$ spectra of emission-line stars that we have obtained using Hectospec, a multi-fibre spectrograph on the 6.5m Multi Mirror Telescope in Arizona. Several additional applications of such software will come online in the near future, such as VPHAS+, one of the three planned European Southern Observatory public surveys which will cover the southern galactic plane in five broad-band colours, or PanSTARRS, a variability survey covering one quarter of the entire sky. Warwick is actively involved in several large survey projects, and the proposed efforts in automated classification will serve a large community of astronomers. The methods developed here are not restricted to the application of astronomical data, but are of general interest for the mining of large data sets in all science disciplines.

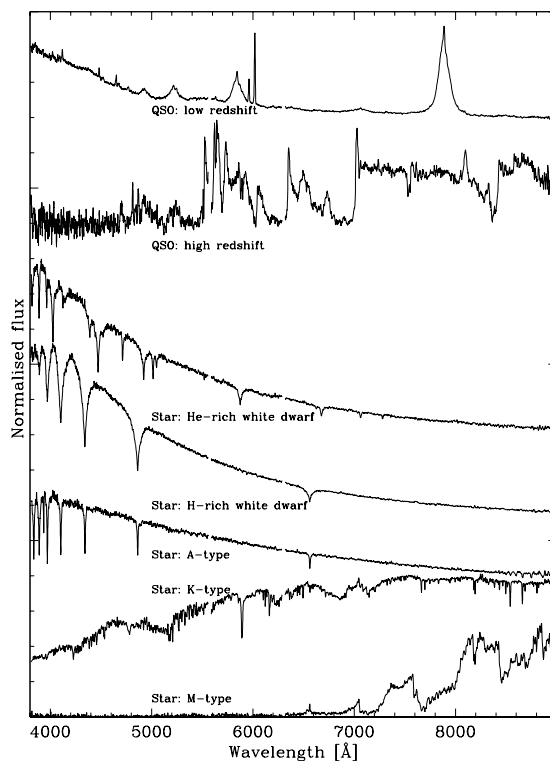


Figure 1: Sample spectra of a small subset of the astronomical object types contained in SDSS. From top to bottom: two quasars at different redshift, two white dwarfs with different atmospheric composition, and three normal stars of different mass.

References

- [1] York, D.G. et al., 2000, AJ 120, 1579
- [2] www.sdss.org
- [3] Gulati, R.K. et al., 1994, ApJ 426, 340
- [4] von Hippel, T. et al., 1994, MNRAS 269, 97
- [5] Rumelhart, D.E. et al., 1986, Nature 323, 533
- [6] Singh, H.P. et al., 1998, MNRAS 295, 312
- [7] Bailer-Jones, C. et al., 1998, MNRAS 298, 361
- [8] Folkes, S.R. et al., 1996, MNRAS 283, 651