Automatic Classification of Subdwarf Spectra using a Neural Network *

C. Winter (cwr@arm.ac.uk) and C. S. Jeffery (csj@arm.ac.uk)
Armagh Observatory, College Hill, Armagh BT61 9DG, N. Ireland

J. S. Drilling (drilling@rouge.phys.lsu.edu)
Dept. of Physics and Astronomy, Louisiana State University, Baton Rouge, LA 70803

Abstract. We apply a multilayer feed-forward back propagation artificial neural network to a sample of 380 subdwarf spectra classified by Drilling et al. (2002), showing that it is possible to use this technique on large sets of spectra and obtain classifications in good agreement with the standard. We briefly investigate the impact of training set size, showing that large training sets do not necessarily perform significantly better than small sets.

Keywords: subdwarf, spectral classification, neural networks

Abbreviations: ANN – Artificial Neural Network

1. Introduction

Stellar spectra require the experience and judgement of a trained expert in order to be classified. However, current and future digital sky survey projects, like the SDSS, along with space-based missions, such as GAIA, will collect huge amounts of spectra – quantities human experts will be unable to cope with. In light of this, investigation into automated classification schemes as supplementary tools is becoming more urgently necessary if we are to stay ahead of the data wave.

Following past examples (Gulati et al. 1994, von Hippel et al. 1994, Bailer-Jones 1996), we aim to establish whether an artificial neural network (ANN) is capable of providing agreeable classifications for a set of subdwarf spectra previously classified by Drilling et al. (2002). Additionally, we briefly investigate how the size and content of the ANN's training data (analagous to spectral classification standards) affects it's ability to provide agreeable classifications.

* This work was carried out as part of the CosmoGrid project, funded under the Programme for Research in Third Level Institutions (PRTLI) administered by the Irish Higher Education Authority under the National Development Plan and with partial support from the European Regional Development Fund.

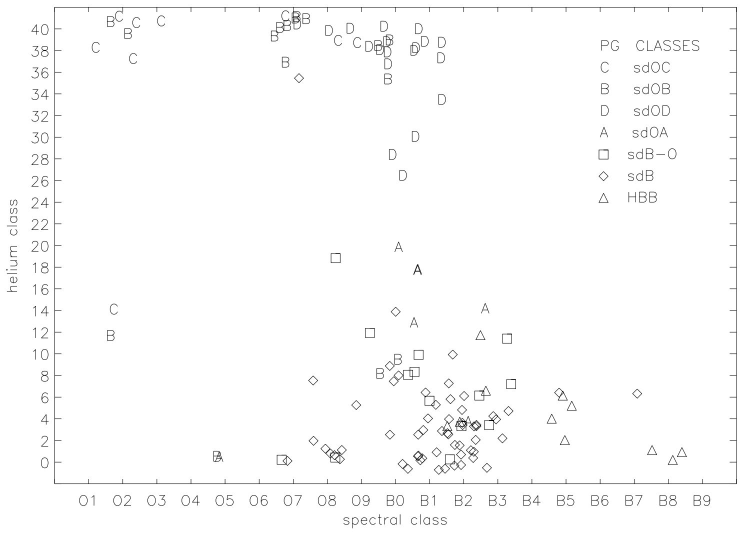

Figure 1. Comparison of Drilling spectral and helium classes with the PG classes (from Drilling et al. 2002)

2. Data Pre-Processing

Our samples of subdwarf spectra were taken from the collection compiled by Drilling et al. (2002) from data provided by Moehler et al. (1990a, 1990b), Dreizler et al. (1990), and Theissen et al. (1993). It comprises a more-or-less representative sample of 174 PG subdwarfs and blue horizontal branch stars, plus a few other stars not included in the PG catalog.

The Drilling classification system uses a spectral type running from sdO1 to sdA (1 – 20), analagous to MK spectral classes. It introduces a helium class (0 – 40) based on H, HeI and HeII line strengths, and uses luminosity classes IV – VIII, where most subdwarfs have luminosity class ∼VII. The mapping between Drilling classes and those used elsewhere, e.g. the PG survey (Green et al. 1986), is illustrated in figure 1. Our data have been been coarsely classified on the helium scale defined by Drilling et al. (2002), with a grain size of 4 helium classes.

Since ANN input data must be in a homogeneous form, any spectra unconformable to a common wavelength range of 4300 – 4850Å, and bin size of 0.6Å, were removed. Crudely rectifying large cosmic spikes, and instrumental end-effects, spectra were then velocity corrected by way of a cross-correlation function. Eliminating spectra with no corresponding spectral classification resulted in a final collection of 380 spectra which were resampled onto a common wavelength range of 4200 – 4900Å, with a bin size of 0.6Å, yielding 1167 data points per spectrum.

3. The Neural Network

An ANN is a statistical pattern recognition algorithm, able to perform a non-linear, parameterised mapping between two domains. Originally inspired by the structure of neuronal cells in the brain, ANNs typically consist of an interlinked, hierarchical structure of processing nodes. See Bishop (1995) for more detailed instruction.

The feed-forward back propagation neural network code STATNET, by Dr. Coryn Bailer-Jones[1], was used in this study.

ANN architecture was kept simple to help determine if spectral classification is within the scope of the algorithm. We used a committee of 5 networks, each consisting of one 1167 node input layer, a hidden layer of 5 nodes, and a single output node.

For each parameter we desired to classify, the data set was divided into two pairs of training and application sets. One pair contained 100 training and 280 application spectra, and the other had 280 training and 100 application spectra. A simple stochastic algorithm was used to select which spectra were allocated to the training and application sets, ensuring the parameter space was evenly represented. An uneven representation limits the ANN's ability to generalise, and reduces performance.

4. Results

ANN classification was limited to spectral type and helium class as ~64% of our spectra are in luminosity class VII, making this class over-represented.

For spectral type, using the training set of 100 spectra, the ANN classified to within 2.09 subtypes, with a correlation coefficient of 0.89. The larger training set of 280 spectra allowed classifications accurate to within 1.99 subtypes, with a correlation coefficient of 0.90.

For helium class, the training set of 100 spectra provided classifications within 4.79 classes, with a correlation coefficient of 0.92. For the larger training set of 280 spectra, classifications were within 4.55 classes, with a correlation coefficient of 0.94. The large errors are due to the coarse grain of the original classification scale.

In each case, we see that the ANN trained on a larger set of 280 spectra yields a classification error not significantly smaller than the ANN trained using a smaller set of 100, suggesting a large training set is not necessarily required for good performance.

[1] http://www.mpia-hd.mpg.de/homes/calj/statnet.html

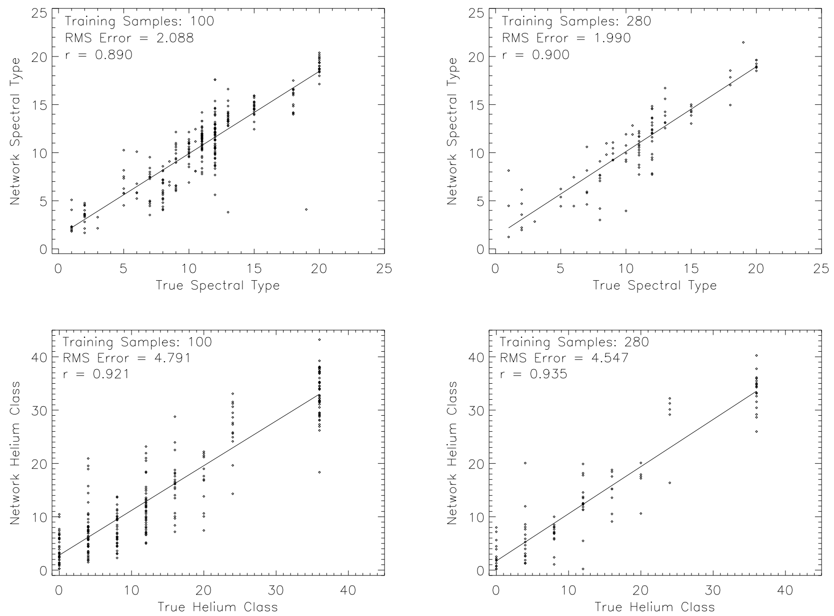

Figure 2. The scatter plots show true classifications against network classifications. Also plotted for each case is a least–squares best fit line.

5. Conclusions

We have established that ANNs are capable of providing spectral classifications agreeable with those made according to the classification standards. In addition, a large training set is not necessarily required for the ANN to yield good results. Future work will allow us to determine if ANNs can yield even better performance.

References

Bailer-Jones C. A. L. 1996, PhD Thesis, University of Cambridge.

Bishop C. M. 1995, *Neural Networks for Pattern Recognition.* (Oxford: Oxford University Press).

Dreizler S., Heber U., Werner K., Moehler S., de Boer K. S. 1990, A&A, 235, 234.

Drilling, J. S., Moehler, S., Jeffery, C. S., Heber, U., and Napiwotzki, R. 2002, *Probing the Personalities of Stars and Galaxies*, ed. Richard Gray, in press.

Green R. F., Schmidt M., Liebert J. 1986, ApJS, 61, 305.

Gulati R. K., Gupta R., Gothoskar P., & Khobragade S. 1994, ApJ, 426, 340.

Moehler S., Richter T., de Boer K. S., Dettmar R. J., Heber U. 1990, A&AS, 86, 53.

Moehler S., Heber U., de Boer K. S. 1990, A&A, 239, 265.

Theissen A., Moehler S., Heber U., de Boer K. S. 1993, A&A, 273, 524.

von Hippel T., Storrie-Lombardi L. J., Storrie-Lombardi M. C., & Irwin M. J. 1994, MNRAS, 269, 97.