SPECTRAL CLASSIFICATION OF GALAXIES

S. Folkes, O. Lahav
Institute of Astronomy, Cambridge, UK

ABSTRACT

The possibility of the parallel acquisition of spectral data alongside the proposed astrometric capabilities of a GAIA type mission will present considerable data analysis problems. To demonstrate one method to handle a large spectral data set, we present an investigation into spectral analysis for the 2dF Galaxy Redshift Survey, which is due to begin this year, utilizing the unique 2dF fibre optic system in conjunction with the Anglo-Australian Telescope. Over the timescale of the project, a library of 250 000 galaxy spectra will be produced. To assess the underlying astrophysics involved in the formation of the spectra, an automatic system is required to spectrally classify the galaxies. We have investigated the use of a Neural Network for such a task, initially for finding the morphology of the galaxies. We present examples of spectral classification using simulated galaxy spectra, from which a measure of the success of the classification method can be derived. We find that at $V = 19.5$ mag over 85% of the galaxies are correctly classified into one of five morphological types.

Keywords: galaxies; classification; spectroscopy; neural networks.

1. INTRODUCTION

With the advent of large astrophysical surveys, (eg. 2dF, Sloan Digital Sky Survey, GAIA) the quantity of spectroscopic data will grow at a rapid pace. Traditionally, experts have analyzed and classified spectra, but this method is both time consuming and subjective. Ideally an automated method is required to firstly reproduce consistently the classification of an expert and secondly to look for new trends in the data which could point towards processes of astrophysical interest. Previous investigations into spectral analysis have been conducted for the classification of QSO spectra (Francis et al. 1992), stellar spectra (von-Hippel et al. 1994), and galaxy spectra (Sodré & Cuevas 1994 and Connolly et al. 1994).

2. THE PROPOSED 2dF GALAXY SURVEY

The 2dF instrument (see Taylor 1994) consists of a new top-end ring for the Anglo-Australian Telescope, incorporating a prime focus corrector lens to give a 2-degree field. A robotic system is used to position 400 fibres across the field, such that the light from the target objects can be sent to two separate spectrographs. In this manner about 400 spectra can be obtained in one observation. The proposed 2dF Galaxy Redshift Survey will attempt to measure 250 000 galaxy redshifts to $b_j = 19.7$, principally for the investigation of large scale structure. Additionally the spectra should produce a good homogeneous data set for spectral analysis, allowing investigations into the distribution of galaxies of varying spectral morphology.

3. SIMULATIONS OF SPECTRA

The Spectrophotometric Atlas of Galaxies (Kennicutt 1991) provides a source of high quality spectra for low redshift galaxies, with morphological classifications. We have taken 38 of these spectra for a range of elliptical and spiral galaxies and shifted them to zero redshift. By considering a sky spectrum and the response function for the 2dF we have simulated spectra for objects in the magnitude range $V = 19.0$ to $V = 22.5$ mag with exposure times of 30 minutes. We have also included a sky subtraction error so that each simulated galaxy spectrum contains between -5 and +5 percent of the sky spectrum. Twenty such simulations, with different random seeds, provides a total data set of 760 spectra at each magnitude.

4. PRINCIPAL COMPONENT ANALYSIS

The technique of Principal Component Analysis can be used to reduce the dimensionality of a set of data (see Murtagh & Heck 1987 for a full description). If the spectrum from a galaxy is represented by the flux at N wavelengths, then the spectrum can be considered as a point in an N dimensional space, with the axes consisting of the fluxes at the different wavelengths. Many such spectra then form clouds of points in the space. Principal Component Analysis finds the vectors in the space along which the data varies most significantly. These vectors are called the Principal Components with the first Principal Component being that which encompasses the most variance in the data. In this way, it is often found that a large amount of the information in the data can be retained by projecting the spectra onto a small number of Principal Components. The simulated spectra can then be projected onto a set of new axes being the principal components determined from the original 38 spectra. It is found that a spectrum can be well reconstructed from the projections onto only a small number of principal components (for example the first eight), and when noisy spectra are being considered it is advantageous to limit the number of components used, since further projections merely reconstruct the noise. For this reason we retain the projections onto the first eight Principal Components, since for the noise levels we are using, this leads to a minimization in the reconstruction error. Fig. 1 shows a galaxy spectrum reconstructed with just eight Principal Components. Fig. 2 shows the same spectrum, with noise added, and its reconstruction from eight Principal Components, indicating how the noise can be reduced.

Table 1 contains the values of this output type for each morphology. The output Type is in the range 0 to 1, since the net considers the galaxies as forming a continuous set.

5. THE NEURAL NET APPROACH

An artificial neural net code is used to classify the spectra. More detailed descriptions of the use of neural nets in astronomy appear in other papers (Hertz et al. 1991, Lahav 1994, Naim et al. 1995). The net we have used has the structure of 8 input nodes, 5 hidden nodes, and one output node. In this case the single output node is used to reflect the spectral morphology of the spectrum being analyzed. The neural net is then trained on 2/3 of the galaxies. This procedure involves a quasi-newton minimization routine (see Hertz et al. 1991), by which the individual weights for the net are varied to minimize the error in the output type. After repeatedly submitting the training set of spectra to the neural net, it is found that the error in the output converges and the training can be considered complete. The remaining 1/3 of the spectra are then presented, and the output type recorded for each spectrum.

6. RESULTS

To assess the performance of the net, the output is binned into the nearest class as defined in table 1. A success rate, given by the percentage of galaxies placed into the correct class, can then be inferred. A much broader classification can also be investigated by performing a two class binning to classify the galaxies as either 'elliptical' or 'spiral'. This entire procedure can be repeated for a data set of galaxies at different magnitudes and the percentage rates plotted against magnitude. Another indication of performance is given by the standard deviation of the output type about the actual type for the galaxies. This gives a true indication of the net performance without the need for artificially binning the results. These results can be seen in Fig. 3. Another interesting procedure, to highlight where misclassifications are taking place, is to form a matrix of the real class against the output class from the neural net. It is found that even at $V = 21$ mag, most of the galaxies have still been classified to within one class (using the five classes defined in Table 1) of their actual morphology.

7. CONCLUSIONS

This investigation indicates that it is feasible to use a neural net to classify the spectra of galaxies and it is clear to see that such methods can also be used to classify stellar spectra, such as those obtained by parallel spectral observations alongside a GAIA type mission. Such a mission would require automated classification methods and this investigation shows that Neural Nets can handle noisy data with additional sky subtraction errors. This analysis also indicates that Principal Component Analysis can be used to both compress data and to reduce noise.

ACKNOWLEDGMENTS

We thank A. Naim, R. Ellis, S. Maddox and the rest of the 2dF Galaxy Survey collaborators for useful discussions concerning spectral classification.

REFERENCES

Connolly, A.J., Szalay, A.S., Bershady, M.A., Kinney, A.L., Calzetti, D. 1994, AJ, submitted

Francis, P., Hewett, P.C., Foltz, C.B., Chafee, F.H. 1992, ApJ, 398, 480

Hertz, J., Krogh, A., Palmer, R.G. 1991, in Introduction to the theory of neural computation, Addison-Wesley, Redwood city, California

Kennicutt, R.C. 1992, ApJS, 79, 255

Lahav, O. 1994, in Vistas in Astronomy, special issue on ANNs in astronomy, 38 (3)

Murtagh, F., Heck, A. 1987, in Multivariate data analysis, Reidel, Dordrecht

Naim, A., Lahav, O., Sodré Jr., L., Storrie-Lombardi, M.C. 1995, MNRAS, submitted

Sodré Jr., L., Cuevas, H. 1994, in Vistas in Astronomy, special issue on ANNs in astronomy, 38 (3)

Taylor, K. 1994, AAO Newsletter, 69

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

Morphology	*Type*
E-S0	0.1
SB0-Sab	0.3
Sb-Sbc	0.5
Sc-Sd	0.7
Sm/Im	0.9

Table 1: Neural Net output for different morphologies

Figure 1: Galaxy spectrum reconstruction from 8 Principal Components.

Figure 2: Noisy Galaxy Spectrum reconstruction from 8 Principal Components.

Figure 3: Neural net performance. a) the percentage success for classifying spectra at different magnitudes for 2 and 5 classification bins. b) The total standard deviation about the correct classification.