Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

Automated Classification of a Large Database of Stellar Spectra

R. K. Gulati and R. Gupta

IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India

P. Gothoskar and S. Khobragade

NCRA, TIFR Center, P.O. Box 3, Pune 411 007, India

Abstract. An Artificial Neural Network (ANN) is a versatile tool which has been used both in academic research and industrial applications. In astronomy, this technique has been used for a variety of applications, such as telescope adaptive optics, classifying galaxies, and separating stars from galaxies. The classification of a large database of stellar spectra, which would be a Herculean task for human classifiers if done visually, is an ideal problem for the ANN technique, which can handle such problems without manual intervention. Recently, increased computational power, combined with improvement in the ANN techniques, has provided an efficient way to perform automatic classification.

We have implemented ANN to classify stellar spectra from large spectral databases. We present here the Multilayer Back Propagation Network (MBPN), which is used to classify stellar spectra obtained in the optical and ultraviolet regions. The performance of MBPN shows that the ANN is capable of classifying ultraviolet stellar spectra to an accuracy of about one spectral subclass for most of the cases. The scope of this technique is expected to be expanded with the availability of large homogeneous digitized stellar spectral databases.

1. Introduction

Progress in ground based and space instrumentation has brought us to a new era of spectroscopy, where a large quantity of good quality stellar spectra has started becoming available through well organized data centers. In order to analyze these spectra and extract useful physical information about stars and stellar systems, we need to develop fast and accurate methods. One way to analyze these spectra is to classify them in terms of common visible properties.

Spectral classification, which conventionally has been done by human classifiers (Houk 1983; Houk & Smith-Moore 1988), involves large, time-consuming efforts. We now require automation of the classification process. The main advantages of automated over human classification is not only the speed with which it can be done, but also accuracy, detection of variability, the elimination of personal error, and the possibility of classification of higher dimensionality.

We (Gulati et al. 1994) have initiated a project to implement automated classification schemes to digitized databases of optical and UV spectra by using conventional metric distance minimization methods and Artificial Neural Networks. We have been using the Multilayer Back Propagation Network (MBPN). Similar efforts have also been employed by another group to classify the stellar spectra of high-dispersion objective prism plates using a neural network scheme (von Hippel et al. 1994). ANN has also been applied to classify a near infra-red database (Weaver 1994).

Here we present the performance of the ANN scheme on libraries of optical and ultraviolet stellar spectra by comparing classifications determined by ANN with those of human classifiers (i.e., catalog classifications).

2. Input and Pre-processing of Optical & UV Data

The optical data were taken from Silva (Silva & Cornell 1992) and Jacoby (Jacoby et al. 1984) libraries. A set of 55 spectra selected from the former library was treated as the template database, and the test database was a set of 158 spectra from the latter library. Both sets were brought to a uniform wavelength range of 3510–6800 Å with 5 Å sampling and 11 Å resolution, and normalized to a value of 100 at 5450Å . Instead of using the full spectral information, a set of 161 wavelength positions was used to monitor the fluxes which are diagnostic of the spectral classes as given by human experts (Jaschek and Jaschek 1990). Catalog classifications of the spectra were taken from the respective libraries. The spectra covered stars of solar metallicity, types O–M, and luminosity classes I–V. Each spectro-luminosity class was coded with a number $x = (1000 \times \mathrm{A1} + 100 \times \mathrm{A2} + (1.5 + 2 \times \mathrm{A3}))$, where A1 was the main spectral type of the star (i.e., O to M types coded from 0.0 to 9.5), A2 was the sub-spectral type (coded from 0.0 to 9.5) and A3 the luminosity class (i.e., classes I to V coded as 0 to 4). For example, a B2I star and a G9.5V star would be coded as 2201.5 and 5959.5, respectively.

The input database for the UV data was the *IUE* Low Resolution Spectra (Heck et al. 1984). A set of 128 spectra spanning 75 spectro-luminosity classes was selected as the template and another set of 83 spectra was used as the test set. The catalog classification was taken from this catalog, where like MK classification, the UV classification is given as O, B, F, etc., as main classes, subclasses ranging from 0.0 to 9.5, and luminosity classes represented as s, g, and d for super-giants, giants, and dwarfs, respectively. The wavelength range of the UV spectra is 1213–3201 Å with 2 Å sampling and 6 Å resolution. The spectra were monitored at 35 wavelength positions which are the diagnostic of these spectral classes as given in Table 1 of the *IUE* catalog (Heck et al. 1984). Here, too, the spectral coding was done by using a number $x = (1000 \times \mathrm{A1} + 100 \times \mathrm{A2} + \mathrm{A3})$, where A1 was the main spectral type of the star (i.e., O to F types coded as 1 to 4), A2 was the sub-spectral type of the star (coded from 0.0 to 9.5), and A3 the luminosity class of the star (i.e., 2, 5, or 8 for s, g, or d). For example, stars dB2.5, gO9.5 and sF7 (ultraviolet classes) were coded as 2258, 1955, and 4702, respectively.

3. The ANN Architectures

We used the standard feedforward supervised neural network, known as "multi-layer backpropagation network (MBPN)" (Rumelhart et al. 1986), for classifying the databases of optical and ultraviolet stellar spectra into different classes of stars. As mentioned earlier, the number of output classes was 55 in the optical case and 75 in the UV case. The input data points for optical and UV data classifiers were different, so the ANN classifiers were selected with different architectures for optical and UV data. The optical classifier was found to be optimal with the configuration 161:64:64:55 and the UV classifier was configured as 35:71:75. The configuration numbers show the input size, hidden nodes, and, at the end, the output nodes, respectively. Once the training was over, the networks could classify the large databases of stellar spectra within a minute, without any human intervention.

4. Performance

The performance of the ANN technique can be judged from Figure 1, which shows the 3D plots of classification errors in luminosity and spectral type on x and y axes and the percentage of total test sample along the z axis, respectively, for optical and UV data. In these plots an ideal classification would appear as a single peak of 100% value in the center of the (x, y) plane, signifying that all spectra are classified correctly with no errors in either luminosity or spectral type. One sub-spectral type error means 100 units error along the y-axis of these 3D plots. Statistical parameters, such as linear correlation coefficients and standard deviations, were computed on the scatter plots (for details see Gulati et al. 1994), and it was found that the classification error for optical was about two subclasses and for UV it is about one subclass, barring a few stars which clearly show more than 100 units of error. These stars require further detailed studies.

5. Conclusions and Future Steps

We do not see any gross mis-classification with the automated ANN scheme and the schemes are quite efficient for large databases. However, we feel that a more homogeneous and complete database is required for the ANN training to perform better. The implementation of ANN on a parallel computer would significantly reduce the training time. In the future, we plan to use a library of synthetic spectra based on stellar atmosphere models (Gulati et al., 1993) to tag information on the stellar physical parameters.

Acknowledgments. R. K. Gulati wishes to acknowledge the generous financial support from the organizing committees of ADASS IV, which allowed him to present this paper at the conference.

References

Gulati, R. K., Malagini, M. L., & Morossi, C. 1993, ApJ, 413, 166

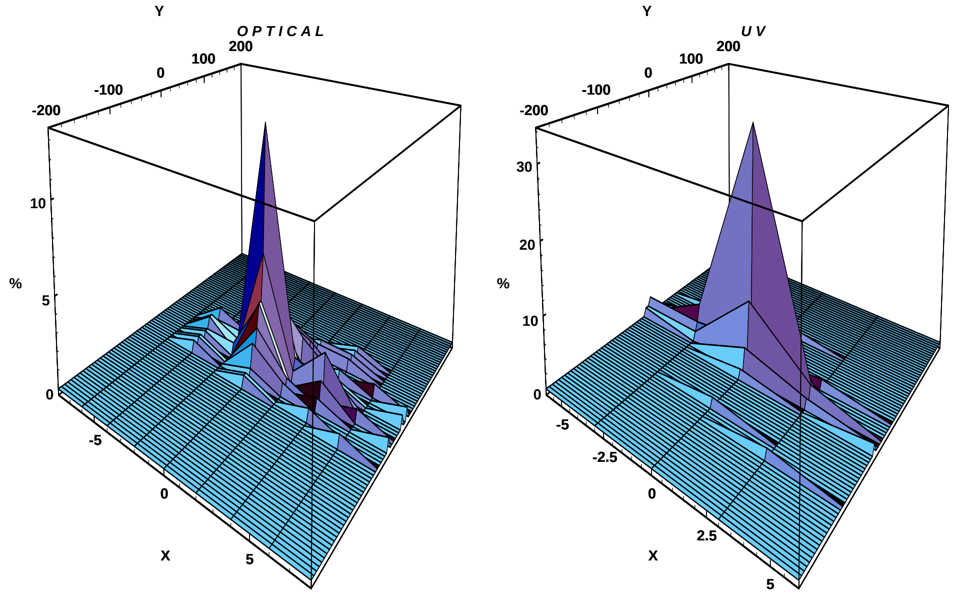

Figure 1. 3D plots for classification errors in luminosity (x-axis) and spectral type (y-axis) vs. the % of total number of test spectra for Optical and UV data.

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

Heck, A., Egret, D., Jaschek, M., & Jaschek, C. 1984, in IUE Low-Resolution Spectra: A Reference Atlas–Part I, Normal Stars, ESA SP-1052

Houk, N. 1983, in The MK Process and Stellar Classification, ed. R. F. Garrison, (Toronto, David Dunlop Observatory), p. 85

Houk, N., & Smith-Moore, M. 1988, in University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars, 1988, Vol. 4

Jacoby, G. H., Hunter, D. A., & Christian, C. A. 1984, ApJS, 56, 257

Jaschek, C., & Jaschek, M. 1990, The Classification of Stars, (Cambridge, Cambridge Univ. Press)

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, Nature, 323, 533

Silva, D. R., & Cornell, M. E. 1992, ApJS, 81, 865

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

Weaver, Wm. B. 1994, in The MK Process at 50 Years: A Powerful tool for Astrophysical Insight, ASP Conf. Series, Vol. 60, eds. C. J. Corbally, R. O. Gray, and R. F. Garrison (San Francisco, ASP), p. 303