Stellar Populations as Building Blocks of Galaxies
Proceedings IAU Symposium No. 241, 2006
A. Vazdekis & R. F. Peletier, eds.

doi:10.1017/S174392130700751X

Cross-checking reliability of some available stellar spectral libraries using artificial neural networks

Ranjan Gupta[1], S. Jotin Singh[2] and Harinder P. Singh[2]

[1]IUCAA, Post Bag 4, Ganeshkhind, Pune 411 007, India

[2]Department of Physics & Astrophysics, University of Delhi, Delhi – 110 007, India

Abstract. Cross-checking the reliability of various stellar spectral databases is an important and desirable exercise. Since number of stars in various databases have no known spectral types and some of the libraries do not have complete coverage resulting in gaps. We use an automated classification scheme based on Artificial Neural Networks (ANN) to cross-classify stars in the Indo-US stellar spectral library (Valdes *et al.* 2004), JHC (Jacoby, Hunter & Christian 1984), ELODIE spectra (Moultaka *et al.* 2004) and STELIB (Le Borgne *et al.* 2003). We have also examined the effects of over-training and over-fitting on the classification efficiency of a Neural Network. It is hoped that such a automated data analysis and validation technique will be useful in the future.

Keywords. Stars, Catalogs, Artificial Neural Networks, Methods: Statistical

1. ANN architecture, Data pipeline and Spectral classifications

Under this scheme there is a training session where the ANN output and the desired output get compared at each iteration and the connection weights get updated till the desired minimum error threshold is reached. The next stage is the testing session where the test patterns are input to the network and output is the classified spectral pattern in terms of the training classes. We have attempted ten combinations of inter-library cross-checks. Table 1 lists these combinations. For example set A1-A3 consists of JHC (with 158 spectra) as the train set and CFLIB (with 1273 spectra), ELODIE (with 1959 spectra) and STELIB (with 247 spectra) as the test sets. In the last column in Table 1,

Table 1. ANN training and test cases. Two hidden layers with 64 nodes each were used for all the cases

Case	Training (Library, No. of Spectra)	Testing (Library, No. of Spectra)	λ Region (used)	Resolution	Class. error (Sub Spectral Type)
A1	JHC, 158	CFLIB, 1273	4100-5500Å	4.5Å	6.7
A2	JHC, 158	ELODIE, 1959	4100-5500Å	4.5Å	5.1
A3	JHC, 158	STELIB, 247	3600-7400Å	4.5Å	8.6
B1	ELODIE, 174	ELODIE, 1959	4000-5500Å	1Å	4.9
B2	ELODIE, 174	ELODIE, 1959	4000-6800Å	1Å	5.7
B3	ELODIE, 174	CFLIB, 1272	4000-5500Å	1Å	7.4
B4	ELODIE, 174	CFLIB, 1273	4000-6800Å	1Å	8.4
C1	STELIB, 247	CFLIB, 1273	4000-6800Å	3Å	6.4
C2	STELIB, 247	CFLIB, 1273	3500-9400Å	3Å	6.7
C3	STELIB, 247	ELODIE, 1959	4000-6800Å	3Å	5.0

Table 2. Various number of hidden layers and corresponding nodes used in the optimization study of ANN.

No. of Hidden Layers	2	4	8
Nodes	16	16	48
per	24	24	64
Hidden	32	32	128
layer	48	48	256
	64	64	
	128	128	
	256	256	

an error of 5 sub-spectral type means that a G2 star is being classified anywhere between F7-G7 and so on.

2. ANN optimization studies

We have also performed an ANN optimization study to see the effects of over-fitting in the training sessions. The train set for this study was JHC spectra with 134 individual classes with 756 fluxes each. The test set was the corresponding train set itself. Table 2 lists out the various networks studied for the determining the optimum number of hidden layer and corresponding hidden nodes. We find that the network configuration of 2 hidden layers with 64 nodes provides the best classification accuracy. Although the larger networks also give comparable accuracy, a 2-64 network is most efficient in terms of computational resources. In addition, for each network 60,000 training iterations were tried to check the classification error and its variation at each iteration. We find that for most of the networks, 30,000 iterations are sufficient else one ends up in a situation of network overtraining.

3. Conclusions

We have used combination pairs from four available libraries, i.e., JHC, ELODIE, STELIB and CFLIB. The results indicate that all these libraries have gaps in their wavelength coverage and also the original catalog spectral classification needs to be checked in greater detail as there are several outliers in the classification scatter plots. Further, there is need to include more spectral and luminosity classes (with more samples per spectral type and luminosity class) and extend the libraries to non-solar type stars.

As an additional study, we also tried various network configurations to obtain an optimization in terms of classification accuracy. This study concerns the over-training and over-fitting in such ANN applications and clearly determines the best network configuration suited for the present application.

References

Le Borgne, J.-F., Bruzual, G., Pello, R., Lancon, A., Rocca-Volmerange, B., Sanahuja, B., Schaerer, D, Soubiran, C. & Vilchez-Gomez, R. 2003, *A&A* 402, 433L (STELIB)

Jacoby, G. H., Hunter, D. A. & Christian, C. A. 1984, *ApJS* 56, 257 (JHC)

Moultaka, J., Ilovaiski, S.A., Prugniel, P. & Soubiran, C. 2004, *PASP* 116, 693 (ELODIE)

Singh, H. P., Gulati, R. K. & Gupta, R. 1998, *MNRAS* 295, 312

Singh, H.P., Yuasa, M., Yamamoto, N. & Gupta, R. 2006, *PASJ* 58, 177

Valdes, F., Gupta, R., Rose, J. A., Singh, H. P. & Bell, D. J. 2004, *ApJS* 152, 251 (CFLIB)