The MK Process at 50 Years
ASP Conference Series, Vol. 60, 1994
C.J. Corbally, R.O. Gray, and R.F. Garrison (eds.)

Automated Classification of Stellar Spectra: Where Are We Now?

T. von Hippel, L. J. Storrie-Lombardi,
M. C. Storrie-Lombardi
Institute of Astronomy, Madingley Rd., Cambridge CB3 OHA, U.K.

M. J. Irwin
Royal Greenwich Obs., Madingley Rd., Cambridge CB3 OEZ, U.K.

Abstract. We briefly review the work of the past decade on automated classification of stellar spectra and discuss techniques which show particular promise. Emphasis is placed on Artificial Neural Network and Principle Component Analysis based techniques, due both to our greater familiarity with these and to their rising popularity. As an example of the abilities of current techniques we report on our automated classification work based on the visual classifications of N. Houk (Michigan Spectral Catalogue, Vol. 1 - 4, 1975, 1978, 1982, 1988).

1. Introduction

The classification of stellar spectra has long been a fundamental tool for determining the physical parameters of bright stars (e.g., Cannon and Pickering 1918-24; Morgan *et al.* 1943). Earlier in this conference Osterbrock (1994), Hoffleit (1994), and McCarthy (1994) reviewed the history of stellar classification, and Mihalas (1994) showed the early achievements and the continuing challenges being met by the conflict of classification with theory. Stellar classification encompasses both the rare objects which refine our understanding of the physics of stellar evolution and stellar atmospheres, and the common objects which we employ towards understanding the general questions of Galactic structure and stellar populations. All of these successes have been built upon the classification work of a few experts – a system which worked well when the data rates where relatively low. Currently, however, very large scale projects are being initiated which will acquire many thousands of spectra per night and which should "reduce" those spectra in nearly real time. The Sloan Digital Sky Survey (Gunn and Knapp 1992), for example, plans on acquiring $\geq 10^6$ spectra over a few year period, though most of these spectra will probably not be stellar. Likewise, a 400 fiber spectrograph is currently being constructed for the Anglo-Australian telescope, making this instrument capable of providing nearly 10^4 fully reduced spectra per night.

In order for spectral classification to leap into this age of very large digital surveys automated techniques must take over much of the time-consuming work done by the human visual classifier. This has been recognized for some

time and considerable effort has been put into automating spectral classification (see Kurtz 1984 for efforts of the previous decades). Some successes have been achieved, though no readily usable and robust techniques have yet emerged. Besides an increase in classification speed, an automated, objective and repeatable technique should also provide a basis for building classification "expertise" at different survey centers.

Techniques which have been developed to automate the classification of stellar spectra can be broken down into two basic categories: criterion-evaluation and pattern-recognition techniques. Criterion-evaluation techniques were tried earlier as they are computationally easier, though they are only a crude representation of the method employed by the human classifier. An early example of criterion evaluation can be found in Jones (1966) where the equivalent widths of a number of hydrogen, calcium, iron, manganese and strontium lines were correlated with spectral type, luminosity class and metallicity. A more modern example of criterion evaluation can be found in Malyuto and Shvelidze (1994, and references therein), where the relative fluxes in a large number of wavelength regions are compared, through stepwise-regression techniques, to the spectral types, luminosities and metallicities. Pattern-recognition techniques have been employed by LaSala and Kurtz (LaSala 1994, and references therein), Weaver (1994) and our group (von Hippel *et al.* 1994). The technique of LaSala and Kurtz is to find the minimum in a weighted vector distance between the program object and the library of standards. Weaver has used supervised artificial neural network (ANN) techniques (1994) for the specific case of Wolf-Rayet stars as well as for the general case of spectral classification. Our work (see also von Hippel *et al.* 1994 for a more detailed look at our techniques) also uses supervised ANN techniques for the general problem of spectral classification. We are creating an expert spectral classification system based on Houk as our expert, and using her original data.

Over the last 20 years, Houk and co-workers (Houk and Cowley 1975; Houk 1978, 1982; Houk and Smith-Moore 1988; Houk 1994) have been creating one of the greatest modern stellar classification libraries by determining the 2- and 3-dimensional spectral types of all stars in the Henry Draper Catalogue (Cannon and Pickering 1918-24, hereafter HD). The 4 catalogues published to date by Houk include over 120,000 stars covering the full range of temperatures and luminosities, and a wide range of metallicities down to the HD magnitude limit of 8.5 to 11, depending on the particular area of the sky. Errors in human classification are hard to quantify, but Houk reclassifies $\approx$ 10% of her stars without knowledge of her earlier classification, and finds an rms error of $\approx$ 0.6 spectral subtypes and $\approx$ 0.25 luminosity types.

Within the spirit of this conference we also note that we believe automating the MK classification scheme is preferable to building a crude model atmosphere spectral evaluation tool since: (1) not enough stars have their astrophysical parameters precisely defined to build the type of bulk classification system we envision; (2) classifications can be mapped into the astrophysical parameters with well-studied relations (e.g., Popper 1980, and references therein); (3) this mapping can be repeated whenever more up-to-date and accurate relations are determined; and, (4) classifications are considered inviolate, since they are based on a series of standards, and therefore should not change with time. The recent advances in plate measuring (e.g., the Automatic Plate Measuring Facility in

Cambridge; Kibblewhite *et al.* 1984, hereafter APM) and the steady advances in computer speed, are also key ingredients in our automation strategy.

Here we examine an expert system based on a back-propagation neural network. We expect neural networks and other pattern-recognition techniques to be free of many of the disadvantages of criterion-evaluation techniques. Line fitting techniques, for example, have the disadvantage of requiring knowledge of the approximate stellar type before they can determine which lines to fit, since very different features will be found at the same wavelengths for different stellar types. Additionally, with criterion evaluation, when different line indices indicate divergent results it is not readily apparent to what degree noise, overall abundances, or abundance patterns are responsible. The most straight-forward pattern-recognition technique is probably cross-correlation, but, in its simplest form, it weights the strongest lines most heavily, though they are not necessarily the features with the highest weight in classification determination. Furthermore, all the classical techniques are based on linear operations. Since we expect to find subtle non-linear relationships between the temperature, surface gravity and metallic-line indicators, a classification scheme which copes with non-linear relationships between parameters ought to offer significant advantages.

2. Data Reduction and Pre-processing

Our data reduction is essentially the conversion of Houk's objective-prism plates and her catalogue into one-dimensional digital spectra tagged with her spectral types and luminosity classes. The "raw" data are $\approx$ 1000 IIaO high-dispersion (108 Å mm^{-1} at Hγ) objective-prism plates taken with the Michigan Curtis Schmidt telescope at Cerro Tololo, Chile (see also Houk 1994). Each plate is 20cm x 20cm in size and at a plate scale of 96.6″/mm subtends an area of $5^\circ \times 5^\circ$ on the sky. We report here the results from the scanning of 6 plates containing 575 HD stars which Houk classified and assigned a quality of 1 or 2 (on a 1 to 4 system). Of these 575 stars 65% are quality 1 and 35% are quality 2. Details of the plate material can be found in Table I, where we list the date each plate was taken, Houk's plate codes, the J2000.0 equatorial coordinates and galactic coordinates of the plate centers, and the number of non-overlapped quality 1 and 2 spectra extracted from each plate.

The APM facility in Cambridge was used to measure the plates. General details of this facility are given in Kibblewhite *et al.* (1984) and the general use of this system for measuring objective prism plates is given in Hewett *et al.* (1985). The scanning procedure used the HST guide-star catalogue as a position driver for the scanner. Digitized scans 18mm x 2mm in area were extracted around each widened spectrum at 22.8μm pixel resolution (corresponding to a 768 x 80 array). After finding and subtracting off the general "sky" background level, the digitized scans were summed orthogonally to the dispersion axis to create one-dimensional, pixel versus density spectra. Following this the extracted spectra were matched against the Houk catalogues, and only those for which Houk has determined a spectral type were kept. In future work we plan to keep all spectra with sufficient S/N for classification, but at present we are just building the expert system, and not classifying previously unclassified stars.

Table 1. Plates Scanned

date	plate	RA	Dec	l	b	num
27/ 5/68	2280	4:13.2	-1:46	194	-35	76
28/ 5/68	2303	3:28.6	1:23	182	-43	99
16/12/68	3316	20:26.9	-0:44	44	-21	96
12/ 1/69	3504	5:29.7	4:43	199	-16	136
15/ 7/69	4825	23:14.6	-4:45	73	-58	73
9/ 8/74	16506	4:35.6	2:58	193	-28	95

The next data reduction step was the differential wavelength calibration of all the spectra within each plate. An objective prism introduces a quadratic distortion of the position of a spectrum on the prism plate compared to a direct plate (in this case the HST Guide Star Catalog) thereby causing the spectrum to shift slightly (up to several pixels) from the expected position. This distortion was mapped as a function of the standard coordinates of the objects by cross-correlating each high-quality spectrum with a high S/N template star of approximately the same type. A quadratic transformation relating direct and prism coordinates was then derived and the solution applied to all stars regardless of S/N. This places all the stars on the same (arbitrary) wavelength system.

The final stage in the pre-processing was to define a stellar "continuum" so that the information in the spectra could be partitioned into line and continuum features. The continuum estimation was accomplished by first running a 51 pixel median filter (Tukey 1971) over each spectrum followed by a 25 pixel boxcar filter. All absorption (or emission) features with scale length less than $\approx$25 pixels are removed by this processing. This is sufficient to track over all the main absorption lines though it underestimates the continuum in areas where the spectral shape changes rapidly, especially through the major molecular features. This will have no effect on our analysis as long as the continuum fits are consistent for a given spectral type and no residual continuum slope remains. We fit the continuum so that we can remove it and thereby produce a classification scheme based on the line information alone, as that is the general method employed by human classifiers. We also examine classification based on the complete spectra. In von Hippel *et al.* (1994) we also study classification based on the continuum information alone. The purpose of studying the continuum-only method was its equivalence to classification based on photometric colors, whereas the difference between the line-only method and the full spectrum method may eventually provide an independent means of determining reddening.

As samples of the spectra used in this work, Figure I presents spectra of 6 stars of various spectral types, along with their continuum fits. From top to bottom the stars are classified by Houk as B9 V, A2 V, F3 V, G3 V, K5 III and M2 III, respectively. The horizontal axis is wavelength in Å, and the vertical axis is emulsion density in arbitrary units. The spectra have S/N $\gg$ 100 and an effective resolution of $\approx$ 2 - 3 Å, resulting in 382 resolution elements over the useful range of 3843 - 5150 Å. Note that for the later spectral types the more limited flux in the blue increases the noise markedly shortward of 4000 Å.

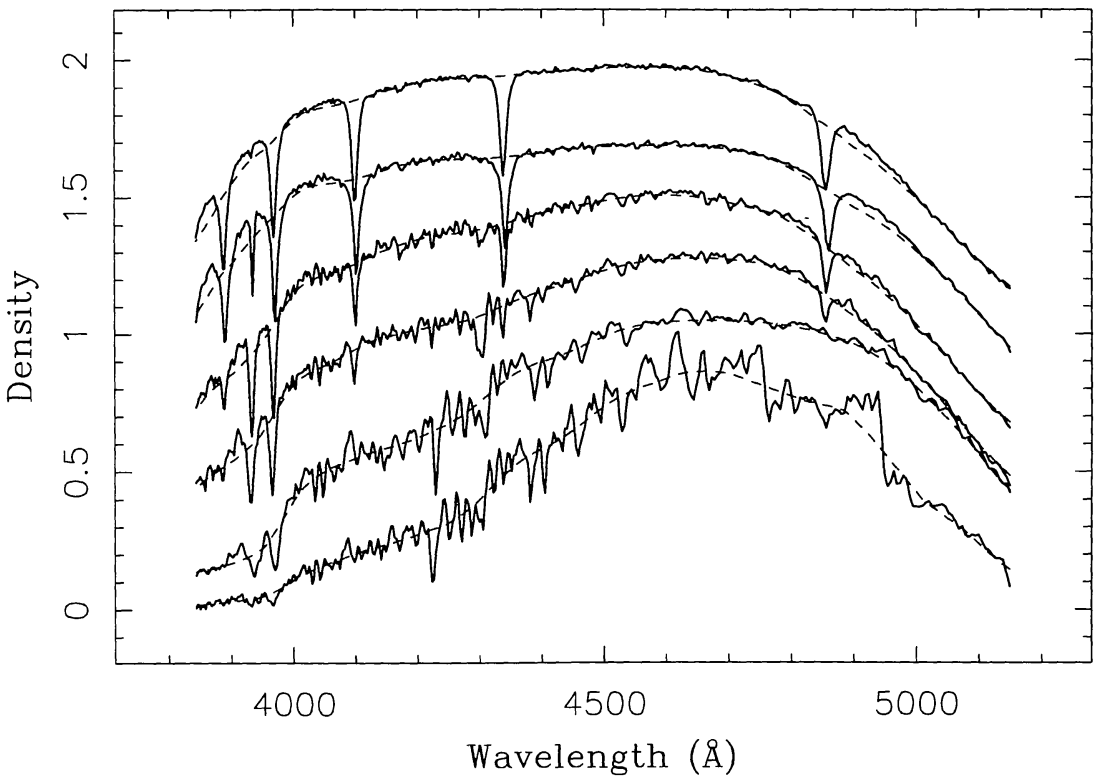

Figure 1. Sample spectra plotted with flux in arbitrary units.

3. Classification Algorithm

3.1. Background

For this study we have utilized an artificial-intelligence technique known as Artificial Neural Networks (ANN) for stellar classification. Originally derived from simplified models of human central nervous system activity (McCullogh and Pitts 1943), ANNs have found utility in signal processing, (Widrow and Winter 1988), adaptive optics (e.g., Merkle 1988), star/galaxy distinction (Odewahn *et al.* 1991), and galaxy classification (Storrie-Lombardi *et al.* 1992). Discovery of an error minimization algorithm for training multiple layers of ANN nodes (Werbos 1974; Parker 1985; Rumelhart *et al.* 1986) has produced considerable interest in a very simple gradient-descent, non-linear minimization ANN technique known as back propagation. Back-propagation ANNs consist of a series of nodes arranged in layers with inputs, summation nodes, outputs, and weighted connections between layers. The classical back-propagation algorithm requires a minimum of three layers: an input layer for bringing the data into the ANN, at least one hidden layer to affect a non-linear mapping from input to output space, and an output layer to provide classification. We have used 383 input nodes, 1 for each of the 382 spectral resolution elements and 1 as a bias node which has a constant value (see Figure II). Weights connecting the bias to the hidden layer and output nodes are modified during training exactly like all other weights. We term any layer between the input and output layers as a hidden layer. We have confined ourselves to the case of a single hidden layer and a single output node

for the purposes of this study to minimize the number of free parameters in our system. The output node is used to provide the temperature classification.

In the standard back-propagation algorithm, the output is defined as a vector in the classification space with an individual node assigned for each class. In our case this would produce a 59-dimension output vector if we considered each temperature classification as a discrete class. The ANN starts with a random initial weight state and guesses at the identity of an incoming target vector by producing an output vector with the value of each output node lying somewhere between 0 and 1. For each star, it then compares its output to the desired classification vector **d** generated by the human expert. For example, we might define $\mathbf{d} = (1, 0, 0, 0, ...)$ for class B0, $\mathbf{d} = (0, 1, 0, 0, 0)$ for class B1, etc.

The comparison is done in terms of a cost function, usually of the form

$$E = \frac{1}{2} \sum_{k} (o_k - d_k)^2 \tag{1}$$

summed over the vector components. Because during training the ANN adjusts its weight space according to the decisions of a human expert, we term it a "supervised" network. This cost function is minimized with respect to the free parameters, the weights w_{ij}, connecting each layer. Minimization over the weights w_{ij} is done using the chain rule (gradient descent). The weights are updated *backwards* from the output layer through the hidden layer(s), and the rate of updating is controlled by the learning coefficient, η. A second coefficient, α (often called the momentum), can dampen the oscillations in minimization sometimes seen as the ANN searches through weight space for an optimal solution. Training set size determines the optimal values for η and α (Eaton and Olivier 1992).

At each node at layer s of the ANN a linear combination over the input x_i from the previous layer $s-1$ is given by $I_j^{(s)} = \sum_i w_{ij}^{(s)} \, x_i^{(s-1)}$. The node fires according to a non-linear sigmoid threshold function usually of the form $f(z) = 1/(1+\exp(-z))$ (in the interval [0,1]) or $f(z) = \tanh(z)$ (in the interval [-1,1]). The ANN functions as a non-linear operator transforming the representation of objects in the input parameter space to the classification space.

After completion of the learning process (i.e. optimizing the weights w_{ij} according to a given training set) the ANN is ready to analyze new data. As we present each input vector the ANN produces an output vector containing its classification coding. In the one output per class configuration the output of an ANN provides a Bayesian *a posteriori* probability for class membership (e.g., Richard and Lippmann 1991; Storrie-Lombardi *et al.* 1992).

3.2. ANN Design

ANN design depends on finding an optimally efficient network architecture for a particular data set. The principal factors include size of the training set, the number of input parameters, number of hidden layer nodes, and number of output nodes. We have maximized the size of the training set (to $\approx$ 500 objects) by training with the data from 5 of the 6 plates, and thereafter testing on the data from the unseen 6th plate. For our output layer we used 1 node since temperature classification is a continuous function rather than a discrete set of

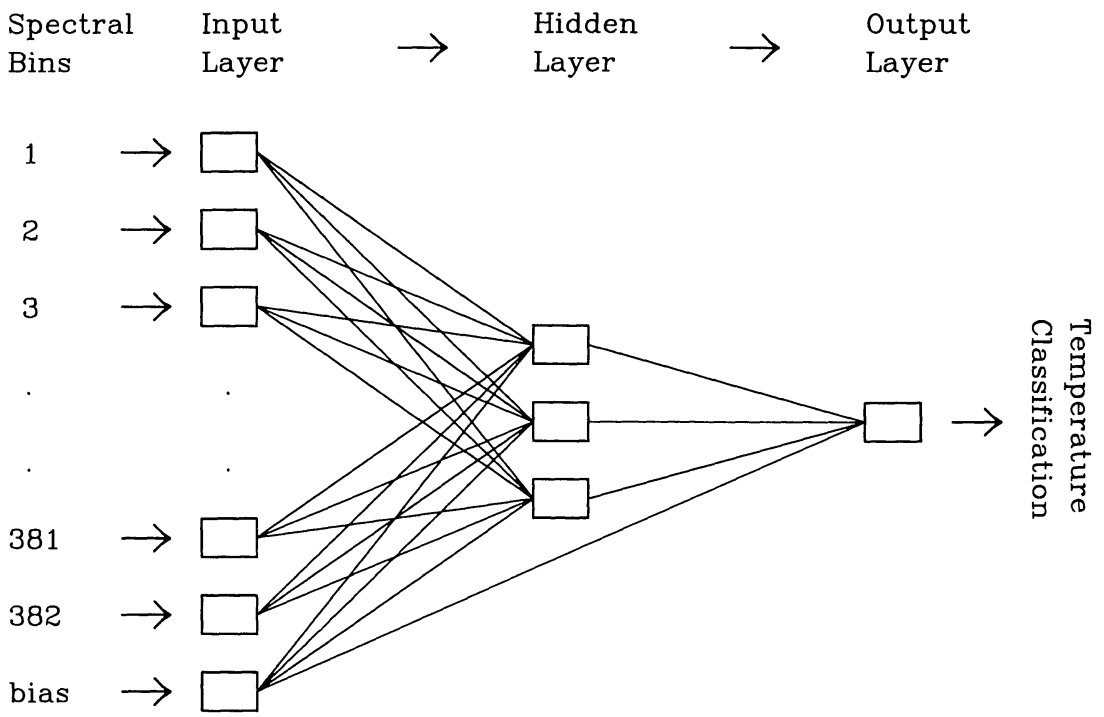

Figure 2. ANN architecture.

output classes. The different spectral type subclasses were converted to floating point numbers from 0 (O3) to 1 (M9), following the 59 subclasses recognized by Houk. This simple linearization of the Houk spectral types mimics the human ability to recognize the effects of small changes in temperature. While we could have constructed an ANN with 59 output nodes, and such an ANN would have had the added advantage of reporting probabilities of class membership, the amount of data presently available to us is not yet sufficient to warrant the greater complications of training and testing such an ANN.

For the present study we have chosen to leave our network fully connected and decrease the total number of weights by decreasing the number of nodes in the hidden and output layers. We have constructed a very simple, minimalist ANN. Back-propagation networks using a single hidden layer train more rapidly and fall into inefficient local minima less frequently than ANNs with two hidden layers (de Villiers and Barnard 1993). We utilize only a single hidden layer containing three nodes. In tests on hidden layer size we have investigated 3, 5, and 7 nodes. The 5 and 7 node cases have nearly the same random and systematic errors as the 3 node case over 1000 training cycles.

At this time we have confined our classification to the temperature domain since we do not have a significant number of stars of different luminosity class at a given temperature class. The spectra described in this series of experiments do not all reside in one luminosity class, but predominantly come from dwarfs at the earlier types and giants at the later types. At this point we have no stars at the classification extremes, either earlier than B3 or later than M4. Figure II depicts our fully connected (each node connects to every node in the next layer) back-propagation network.

3.3. ANN Training

Once the basic ANN architecture has been established, values for the learning (η) and momentum (α) coefficients must be fixed. The best values for η and α seem to be best determined by experience with the data. Poor values for these coefficients cause slow or incomplete learning. We have found that for our current data set $\alpha = 0.9$ and $\eta = 0.01$ produce stable temperature classifications.

Training the ANN can take hundreds or even thousands of iterations. The major portion of the error minimization occurs in the first few dozen iterations, followed by an exponential decrease in the rate of learning. Reasonable criteria for stopping learning include: (1) the rate of learning has effectively approached zero, and (2) a comparison of ANN performance with its training and unseen testing sets, indicates good generalization to the testing set without memorization of the training set. We have found that our networks reach a stable state and show good generalization after 1000 passes through our training samples.

4. Results

In order to maximize the number of stars used to train the ANNs we trained our ANN on the spectra from 5 plates and tested on the spectra from the remaining 1 plate. At present, with $\approx$ 500 training spectra containing 382 independent data points (resolution elements) and an ANN architecture with 383 (= 382 input nodes + 1 bias node) $\times$ 3 (hidden layer nodes) + 4 (= 3 hidden layer nodes + 1 bias node) free parameters (weights), the final ANN weights are overdetermined by a factor of $\approx$ 150.

A persistent problem at present is the large number of objects contaminated by overlapping spectra. For the present these objects were identified and removed by eye. Undoubtedly some spectra remain which are affected by overlapping objects and this may degrade the quality of the classification results.

Figure III shows the result of a typical training and testing session for our ANNs. Panel 3a is the training case and panel 3b is the testing case. The horizontal axis for both panels is the catalogue classification of Houk and the vertical axis is the ANN classification. A line of slope unity is plotted in each panel to allow the eye to assess the quality of the ANN determinations. The self-test plots are created by passing the training set back through the ANN, essentially showing how well the ANN reached the global minimum of the solution space during training. The testing plots are created by passing the previously unseen data through the trained ANN. The testing data will generally have a larger scatter than the self-test data due to the imperfect ability of the ANN to generalize. Nonetheless, in all cases with these data, the test data have only slightly more scatter than the training data, and high quality classifications are achieved.

The quality of the classifications can be seen quantitatively from the systematic and random errors of the training and testing processes. For the line-only case the systematic classification offsets for the training data was 0.19 spectral subtypes, in the sense of the Houk value - the ANN value, (e.g., 1.0 is the difference between an F6 and an F7 star). The systematic classification offset for the testing data was 0.24 spectral subtypes. The classification uncertainties (root-mean-square differences of Houk - ANN) are 1.39 and 2.15 spectral subtypes for

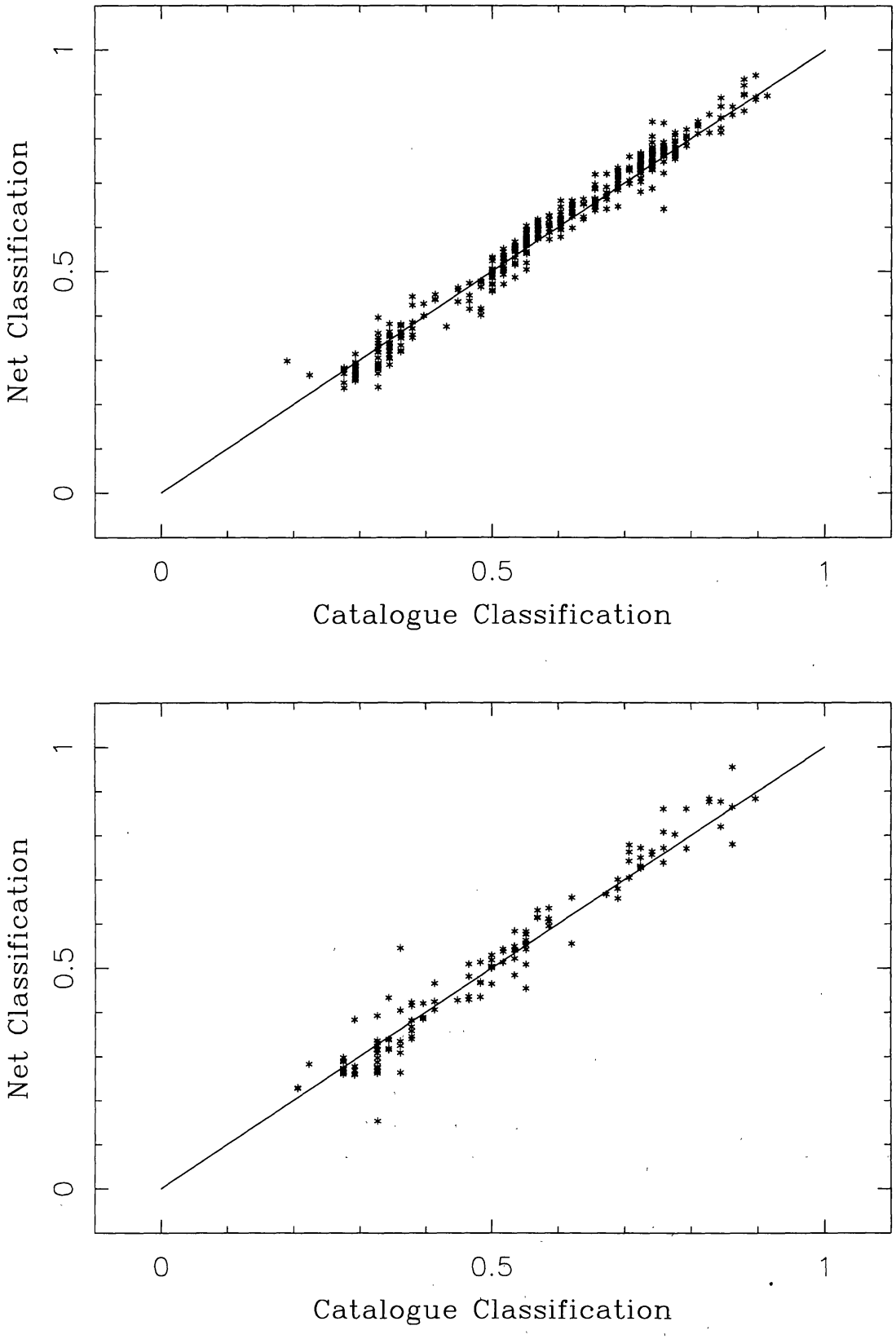

Figure 3. ANN results for typical training (Panel A, above) and testing (Panel B, below) sessions. The horizontal axes are the Houk classifications and the vertical axes are the ANN classifications. Unit slope lines are overplotted.

training and testing, respectively. Since the distribution of errors was somewhat non-normal for the line-only case we also computed the distance from the mean which contained 68% of the classification differences. Thus 68% of the training spectra are within 1.18 spectral subtypes of the mean and 68% of the testing spectra are within 1.72 spectral subtypes of the mean. The complete spectra (line + continuum) case had somewhat lower uncertainties and also low systematics. The complete spectra systematic classification offsets were 0.59 and 0.64 subtypes for training and testing, respectively. The classification rms uncertainties were 1.38 and 1.72 subtypes for training and testing, respectively. Thus our ANNs are currently able to classify stars from B3 to M4 with and rms error of $\approx$ 1.7 spectral subtypes.

Figure III and the above statistics also show that no stars are grossly misclassified. Approximately 5% of the spectra are still contaminated with overlapping stars, however, and this probably degrades the quality of the ANN results. We see systematic errors of $\approx$ 2 spectral subtypes at the blue and/or red ends for some of the training and testing sets. The systematics for the bluest stars (B3-B5) are the result of the subtle features which cause the early B stars to differ from the later B stars, and the lack of examples in this temperature region. The systematics for the reddest stars (M3-M4) are the result of the diminishing amount of distinguishing spectral information available in blue spectra of very late type stars, as well as the limited number of examples of these objects.

Our classifications for the line-only case show only a small overall systematic offset but a somewhat high classification rms uncertainty (σ = 2.2). The classification uncertainty is elevated by the $\approx$ 5% of the classifications which are furthest removed from their nominal catalogue value, leading to the non-normal error distribution. These objects had a large amount of noise somewhere in their spectra, generally in the blue. A further $\approx$ 10% of the stars had considerable noise in the blue and were still properly classified. This highlights the degree to which the ANNs were generally robust to noise.

As an alternate method of transforming spectra into classifications we used the data compression property of Principal Component Analysis (PCA) to reduce our spectra to 20 principal components. We found that the first PCA component accounted for 71% of the spectral variation, the first 2 components 84%, the first 5 components 94%, the first 10 components 97%, and the first 20 components 99%. The rest of the spectral variation seemed to be noise and the contaminating effects of overlapping spectra. Use of PCA had the advantage of removing much of the contaminating effects of overlapping spectra.

We combined the 20 PCA components for all of our stars with the Houk classifications, and trained and tested a smaller ANN. Use of the PCA components had the advantage of allowing us to build an ANN with only 21 input nodes (20 PCA inputs + bias), nearly 20 times fewer input nodes than our standard ANN. We thus were able to train these ANNs 20 times faster. This method also allowed us to experiment with PCA as a form of data compression and to compare different principal components to the classifications. We found no simpler correlations between the principal components and the classifications than previously have been recognized in the criterion evaluation work of other authors. Our ANNs with PCA inputs produced nearly identical answers to those ANNs which used the full spectral information, at least for temperature classifications. We are still actively investigating the full utility of PCA preprocessing.

5. Conclusions

We have presented initial automated stellar spectral classifications based on 575 stars scanned from 6 objective prism plates at the APM Facility. We have employed a simple back-propagation artificial neural network which is trained to duplicate the classifications of Houk (1975, 1978, 1982, 1988). We find that even with the limited data set now in hand we can determine decent quality temperature classifications for spectral subtypes B3 to M4, though we do not yet have the ability to determine luminosity classifications. We classify spectra to within 2.2 spectral subtypes based on the spectral line information alone and to within 1.7 spectral subtypes based on the full spectral information.

Future work will be based on more plate material so that we can increase our classification precision and extend our techniques to the determination of luminosity classes and, eventually, stellar abundances. We shall also explore alternative classification techniques, including unsupervised ANNs. The supervised networks presented here serve as error minimization algorithms, replicating the choices and decision patterns of their trainer. Unsupervised ANNs, however, are capable of deciding on an appropriate classification scheme without benefit of a trainer. Such unsupervised ANNs can learn to group or cluster objects into bins, usually by determining the Euclidean (or any other appropriate metric) distance between individual vectors. These ANNs operate much like classical vector analysis strategies, and they may allow us to identify previously unrecognized stellar classifications.

We believe our results indicate the potential precision, repeatability and objectivity of automated stellar spectral classification. Furthermore, only automated techniques have the speed to extend the expertise of the human classifier to the vast amount of data which should be produced by the next generation of spectral survey instruments.

Acknowledgments. It is a pleasure to thank Nancy Houk for her considerable efforts in helping us begin this program and for loaning us the plate material. We had assistance with numerical techniques from Geraint Lewis and Derek Richardson. We thank Ofer Lahav for very useful discussions.

References

Cannon, A.J., Pickering, E.C. 1918-1924, *The Henry Draper Catalogue*, Harvard Annals, **91-99**

de Villiers, J., Barnard, E. 1993, *IEEE Trans. Neural Networks*, **4**, 136

Eaton, H.A.C., Olivier, T.L. 1992, *Neural Networks*, **5**, 283

Gunn, J.E., Knapp, G.R. 1992, *Astronomical Surveys*

Hewett, P.C., Irwin, M.J., Bunclark, P.S., Bridgeland, M.T., Kibblewhite, E.J., He, X.T., Smith, M.G. 1985, MNRAS, **213**, 971

Hoffleit, D. 1994, this volume

Houk, N. 1994, this volume

Houk, N., Cowley, A.P. 1975, *University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars*, Vol. 1

Houk, N. 1978, *University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars*, Vol. 2

Houk, N. 1982, *University of Michigan Catalogue of Two-Dimensional Spectral Types for the HD Stars*, Vol. 3

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

Jones, D.H.P. 1966, *Roy. Obs. Bull.*, **126**, 219

Kibblewhite, E.J., Bridgeland, M.T., Bunclark, P.S., Irwin, M.J. 1984, *Astronomical Microdensitometry Conference*, ed. D.A. Klinglesmith, Washington D.C.: NASA, 277

Kurtz, M.J. 1984, *The MK Process and Stellar Classification*, ed, B.F. Garrison, Toronto: David Dunlop Observatory, 136

LaSala, J. 1994, this volume

Malyuto, V., Shvelidze, T. 1994, this volume

McCarthy, M.F. 1994, this volume

McCullogh, W.S., Pitts, W.H. 1943, *Bull. Math. Biophysics*, **5**, 115

Merkle, F. 1988, *Messenger*, **52**, 5

Mihalas, D.M. 1994, this volume

Morgan, W.W., Keenan, P.C., Kellman, E. 1943, *An Atlas of Stellar Spectra*, Chicago: University of Chicago Press

Odewahn, E.B., Stockwell, R.L., Pennington, R.M., Humphreys, R.M., Zumach, W. 1991, AJ, **103**, 318

Osterbrock, D.E. 1994, this volume

Parker, D.B. 1985, *Learning-logic*, Report TR-47, Cambridge, MA: MIT Press

Popper, D.M. 1980, ARA&A, **18**, 115

Richard, M.D., Lippmann, R.P. 1991, *Neural Computing*, **3**, 461

Rumelhart, D.E., Hinton, G.E., Williams, R.J. 1986, *Parallel Distributed Processing*, eds. J.L. McClelland, D.E. Rumelhart, and the PDP Research Group, Cambridge, MA: MIT Press, Vol. 1

Storrie-Lombardi, M., Lahav, O., Sodré, L., Storrie-Lombardi, L. 1992, *MNRAS*, **259**, 8p

Tukey, J. W. 1971, *Exploratory Data Analysis*, Reading: Addison-Wesley

von Hippel, T., Storrie-Lombardi, L.J., Storrie-Lombardi, M., and Irwin, M. 1993, *MNRAS*, (submitted)

Weaver, Wm.B. 1994, this volume

Werbos, P.J. 1974, Ph.D. Thesis, Harvard University

Widrow, B., Winter, R. 1988, *Computer*, **21**, 25

Discussion

Gray: Have you gone back into your network and examined the weightings that the network has come up with after training? Is it classifying the lines, the continuum or some ill-defined combination of the two?

von Hippel: We have looked at this in a cursory way. Our tests with three nodes in the hidden layer indicate that one node is generally looking at all the major features, another is looking at Ca H&K and Hδ and the third at the continuum shape (if given that information).

Lloyd Evans: The IRAS catalogue has revealed new types of spectra with magnitudes 12-15. More must remain to be discovered but lack IR excesses to pinpoint them.

von Hippel: If we find new types of stars, we will require calibration by humans. Houk and Garrison have indicated their willingness to help us in this regard.

Lloyd Evans: This technique can only carry conviction if sample results are subjected to blind crosscheck by experienced classifiers. Automated variable star surveys – a simpler problem – have produced some very inaccurate results because there were inadequate human checks.

von Hippel: I showed you a "blind crosscheck" in the testing sample. These spectra had not been seen by the network and were classified to within ~ 1.5 spectral subtypes.

Rountree: Is there a systematic effect in the differences between the neural net classification and Houk classification?

von Hippel: Yes, at early and late types there are small systematic departures between Houk's types and the neural net types. This is due to the few examples at the temperature extremes and to the small amount of information in the blue for the late types, which has also given human classifiers difficulty.

Rountree: In the architecture of your neural net, is there one input node that by-passes the hidden layer?

von Hippel: Yes, this is the "bias" node (constant input signal = 1) which gives the neural net an additional free parameter for each hidden and output node, allowing these nodes to use the most appropriate part of the sigmoid function.

Drilling: Have you devised tests which shows how well your procedure will detect peculiarities?

von Hippel: We have not dealt with peculiarities yet, but intend to try unsupervised neural networks for this purpose. Such networks may even identify new groups of peculiar stars.

La Sala: How do you remove the continuum residual? Continuum effects can contain as much power as the lines and can bias the results.

von Hippel: We remove the continuum by a combination of linear and median boxcar smoothing, then dividing this "continuum" out.

La Sala: Have you dealt with reddened spectra?

von Hippel: Not yet.

La Sala: Do you need to retrain your system for each new instrument or instrumental set-up? How many spectra are needed to train?

von Hippel: Probably. As long as the flux scale can be accurately remapped a previously trained neural network could be used. Houk's plate data, however,

probably cannot be precisely flux calibrated. The number of spectra needed to retrain depends on the accuracy and the spectral type–luminosity domain which is required, but I don't yet know the actual figure.

Craine: I would like to comment that the efforts under discussion in this session have very close analogs in medicine related to comparison of various medical systems, particularly those involving, at some point, a human expert. A very useful tool has been developed from signal-detection theory, known as the Receiver Operating Characteristic (or ROC analysis). This tool is unknown in astronomy and has considerable potential power. The classical reference is: Swets, J. A. and Pickett, R. M., *Evaluation of Diagnostic Systems*, Academic Press: New York (1982). Examples of applications can be found in Craine, E. R. and Seely, G. W. *SPIE Infrared Technology*, **572**, 15 (1985) and Craine, E. R. and Cramer, B. L., *SPIE Medical Imaging*, **1444**, 389 (1991).

von Hippel: I appreciate the comparison with medicine and the references. My collaborator, Michael Storrie-Lombardi, is a medical doctor and his interests lie in this domain as well.

Corbally: Could you explain why the quality and density of the data do not have to be uniform? In contrast, for visual classification both quality and exposure (or S/N) *are* important for learning and accuracy.

von Hippel: The library of training spectra may have an uneven distribution of types and quality as the neural network generalizes from all the spectra it has seen. By density I did not mean plate density, but rather the density of types at any given location in the stellar library.