Automatic Determination of Stellar Atmospheric Parameters Using Neural Networks and Instance-Based Machine Learning

Olac Fuentes
Instituto Nacional de Astrofísica, Óptica y Electrónica
Luis Enrique Erro # 1
Santa María Tonantzintla, Puebla, 72840, México

November 7, 2001

Abstract. In this paper we show how machine learning methods can be effectively applied to the problem of automatically predicting stellar atmospheric parameters from spectral information, a very important problem in stellar astronomy. We apply feedforward neural networks, Kohonen's self-organizing maps and locally-weighted regression to predict the stellar atmospheric parameters effective temperature, surface gravity and metallicity from spectral indices. Our experimental results show that the three methods are capable of predicting the parameters with very good accuracy. Locally weighted regression gives slightly better results than the other methods using the original dataset as input, while self-organizing maps outperform the other methods when signi£cant amounts of noise are added. We also implemented a heterogeneous ensemble of predictors, combining the results given by the three algorithms. This ensemble yields better results than any of the three algorithms alone, using both the original and the noisy data.

Keywords: data analysis, stellar atmospheric parameters, neural networks, instance-based machine learning, ensembles

1. Introduction

In astronomy and other scienti£c disciplines, we are currently facing a massive data over¤ow. With the development of new sophisticated detectors that cover virtually all of the electromagnetic spectrum, and with the rapid development of the Internet, astronomers have at their disposal amounts of information that are vastly larger than what they can analyze by traditional manual or semimanual means.

In recent years astronomers and arti£cial intelligence researchers have started collaborating towards the goal of automating the task of analyzing astronomical data. By far the most commonly used methodology has been feedforward neural networks, generally applying the backpropagation algorithm for training. In galactic astronomy they have been applied to morphological classi£cation of galaxies (Storrie-Lombardi et al., 1992), star/galaxy separation (Odewahn and Nielsen, 1994) and spectral classi£cation of galaxies (Connolly et al., 1995; Folkes et al., 1996). In stellar astronomy, they have been applied to classi£cation of stars (Storrie-Lombardi et al., 1994; Gulati et

al., 1994; Weaver and Torres-Dodgen, 1995) and to physical parameterization of spectra (Bailer-Jones et al., 1997).

While arti£cial neural networks have had remarkably good results in some problem domains, it is worthwhile to investigate if other automated methods can provide more accurate results. In this paper we compare the performance of feedforward neural networks, the most commonly used method in automated astronomical prediction and classi£cation, to two other methods, Kohonen self-organizing maps (Kohonen, 1988) and locally-weighted regression (Bouttou and Vapnik, 1992; Atkeson et al., 1997). We also implemented a heterogeneous ensemble of predictors, combining the results given by the three algorithms.

The organization of the remainder of this paper is as follows: Section 2 describes the data used in our experiments, Section 3 describes the learning algorithms we applied, Section 4 presents experimental results and Section 5 presents conclusions and directions for future research.

2. Data

In our experimental work we used a dataset due to Jones (1996). It is a homogeneous catalog consisting of 48 spectral indices for 684 stars observed at KPNO with the coudé feed instrument that has been made available by the author at an anonymous ftp site at the National Optical Astronomy Observatories (NOAO)

The spectral indices were measured from the spectra in the wavelength regions [3820 Å- 4500 Å] and [4780 Å- 5450 Å] by following the de£nition of the Lick indices (Worthey et al., 1994), the Rose indices (Rose,1994) and new Lick-type Balmer indices (Jones and Worthey, 1995). The wavelength region [3820 Å- 4500 Å] has the following Lick indices: H_δ, CN_1, CN_2, Ca4227, G4300, H_γ and Fe 4383 both measured in the form of equivalent width and magnitude. Also, the region contains the following Rose indices: H_δ/4045, H_δ/4063, SrII/4045, SrII/4063, pGband, Hg/4325, 384/4353, p[Fe/H],CaII, p4220/p4209, eqwCaI, eqw 4045, eqw CaIIK, eqwCaIIH, eqwH_δ. From the region [4780 Å- 5450 Å], Jones has derived the following Lick indices: H_β, Fe5015, Mg_1, Mg_2, Mg_b, Fe5270, Fe5335, and Fe5406 both in equivalent width and magnitude units. As far as the atmospheric parameters in the catalog, there are two sets; one is taken directly from the literature and the other is derived after applying systematic corrections. In this paper we use the later one, assuming that it is a true representative of the stars' observational properties.

Out of the total number of stars in the catalog, we selected those that satisfy the following criteria: each star should have all three of the physical parameters: effective temperature (Teff), surface gravity (log g) and logarith-

mic difference between the relative abundance of iron in the star and the relative abundance of iron in a standard star ([Fe/H]), listed in the catalog. Also, stars in the temperature range between 3000-12000 K were selected because there are very few stars outside of this range. A total of 651 stars satisfy these criteria.

3. Methods

3.1. Feedforward Neural Networks and Self-Organizing Maps

It is beyond the scope of this paper to give a detailed description of feedforward neural networks and self-organizing maps. Instead, we will brie¤y describe how they were adapted to our task. For an introduction to neural networks we refer the reader to any of the excellent books that have been published in recent years, such as (Hecht-Nielsen, 1990) and (Bishop, 1996).

We applied standard feedforward neural networks without any major modi£cation. We used the backpropagation algorithm for training and the tan-sigmoid function as a transfer function. The self-organizing map algorithm was modi£ed to perform in a supervised fashion by combining into each training example and prototype associated with each unit the input and output parameters. That is, each training example (and prototype) is represented by a vector of length 51, where the £rst 48 elements correspond to the 48 spectral indices and the £nal 3 elements correspond to the output parameters (the stellar atmospheric parameters). After training, we extract the prototypes in each unit and build a table, where each entry maps the input parameters to the output parameters contained in the same prototype. We then use this table to predict the output parameters of a given star by £nding its 3-nearest neighbors in the input space and assigning to it the output parameters given by the weighted average of the output parameters of its neighbors, where the weight is inversely proportional to the Euclidean distance in the input space between that star and its neighbors.

3.2. Locally-Weighted Regression

Locally-weighted regression belongs to the family of instance-based learning algorithms. In contrast to most other learning algorithms, which use their training examples to construct explicit global representations of the target function, instance-based learning algorithms simply store some or all of the training examples and postpone any generalization effort until a new instance must be classi£ed. They can thus build query-speci£c local models, which attempt to £t the training examples only in a region around the query point.

In this work we use a linear model around the query point to approximate the target function.

Given a query point $\mathbf{x}_q$, to predict its output parameters $\mathbf{y}_q$, we assign to each example in the training set a weight given by the inverse of the distance from the training point to the query point: $w_i = \frac{1}{|\mathbf{x}_q - \mathbf{x}_i|}$. Let W, the weight matrix, be a diagonal matrix with entries $w_1, \ldots, w_n$. Let X be a matrix whose rows are the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the input parameters of the examples in the training set, with the addition of a "1" in the last column. Let Y be a matrix whose rows are the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$, the output parameters of the examples in the training set. Then the weighted training data are given by $Z = WX$ and the weighted target function is $V = WY$. Then we use the estimator for the target function $\mathbf{y}_q = \mathbf{x}_q^T (Z^T Z)^{-1} Z^T V$.

Thus, locally weighted linear regression is very similar to least-squares linear regression, except that the error terms used to derive the best linear approximation are weighted by the inverse of their distance to the query point. Intuitively, this yields much more accurate results than standard linear regression because the assumption that the target function is linear will not hold in general, but is a very good approximation when only a small neighborhood is considered.

3.3. ENSEMBLE METHODS

An ensemble of classi£ers is a set of classi£ers whose individual decisions are combined in some way, normally by voting. Somewhat surprisingly, ensembles often yield better results than the best of the individual classi£ers that make them up. Ensembles work only if the individual classi£ers are more accurate than random guessing and their errors are not strongly correlated. Several methods have been proposed for constructing ensembles, including bagging (Breiman, 1996), boosting (Freund and Shapire, 1996), error-correcting output-coding (Dietterich and Bakiri, 1995) and randomization (Kwok and Carter, 1990; Ali and Pazzani, 1996). Normally, ensembles are homogeneous in the sense that the same learning algorithm is implemented by each member of the ensemble, and they are forced to produce noncorrelated results by proving slightly different training sets for each member, as in boosting and bagging, by manipulating the outputs, as in ECOC, or by slightly modifying some parameters of the algorithms, as in randomization. Here we propose a heterogeneous ensemble that combines, by means of unweighted average, the results given by the three learning algorithms mentioned previously. It will be shown that this heterogeneous ensemble is more accurate that any of the three algorithms taken individually.

4. Experimental Results

We implemented locally-weighted regression and self-organizing maps in Matlab^{TM}, and used the feedforward neural network implementation that comes with the Matlab^{TM} Neural Network Toolbox. For all the experiments reported we used £ve-fold cross-validation, that is, we partitioned the original dataset randomly into £ve equally-sized subsets and performed £ve experiments, using in each experiment one of the subsets for testing and the other four for training.

Before performing the experiments, the indices in the dataset were standardized so as to have a zero mean and unit variance across the dataset. This eliminates the potential problem of the dependence of the results on the particular units in which different indices were measured, but it also implicitly assumes that all the indices are equally important for each prediction task.

Figure 1 plots the catalog versus predicted parameters after applying locally weighted regression to the prediction of the parameters. To build the linear model around the query point we chose the 200 closest points and weighted each point by a factor inversely proportional to its Euclidean distance to the query point. Figure 2 shows the results of applying a self-organizing map. We used a 15 by 15 two-dimensional grid, with a total of 225 units and we trained it for 25 epochs. Figure 3 shows the results obtained applying a feedforward neural network to the prediction of the parameters. The network had a con£guration of 48 input units, 8 hidden units and three output units, one for each of the three parameters mentioned at the end of Section 2. We trained it for 200 epochs using the backpropagation algorithm. Figure 4 shows the results of applying the heterogeneous ensemble method, which combines the results of the three other methods by means of a simple (unweighted) average. It can be seen that for all four methods, temperature can be predicted with the highest degree of accuracy, while metallicity is harder and surface gravity is still more dif£cult. This is consistent with results attained by human experts. Table I shows the mean absolute errors for each of the methods and each of the parameters. This was obtained by averaging the results of three runs of £ve-fold cross-validation for each method. It can be seen that LWR gives the best performance among the individual methods, and that the ensemble method gives the best overall results.

For a second experiment we added noise to the original data to test the noise-sensitivity of each of the methods. Table II shows the results obtained after adding Gaussian noise with zero mean and unit standard deviation to each of the attributes of the original data. All methods suffer a degradation in performance, as expected, but surprisingly LWR is not noticeably more affected by noise than the neural network methods. This may be due to the relative large number of neighboring points we chose to build the local approximation, and the fact that the noise, being white, tended to cancel across

Table I. Mean absolute errors in the prediction of stellar atmospheric parameters using the original data.

Algorithm	Teff [K]	Log g [dex]	[Fe/H] [dex]
LWR	101.25	0.209	0.111
SOM	135.08	0.240	0.134
FFNN	127.28	0.206	0.123
Ensemble	97.80	0.192	0.107

Table II. Mean absolute errors in the prediction of stellar atmospheric parameters using the three learning algorithms and adding Gaussian noise with zero mean and $\sigma = 1$ to the original data.

Algorithm	Teff [K]	Log g [dex]	[Fe/H] [dex]
LWR	279.81	0.548	0.259
SOM	257.62	0.497	0.276
FFNN	303.67	0.519	0.315
Ensemble	252.56	0.468	0.259

the neighboring points. Both LWR and SOM still perform slightly better than FFNN, and the ensemble method still yields better results than any of the individual methods.

5. Conclusions and Future Work

We have performed an experimental comparison of locally weighted regression, self-organizing maps and feedforward neural networks, three well-known machine learning methods, applied to the problem of predicting stellar atmospheric parameters from spectral indices. Our experimental results show that all three methods are capable of yielding predictions that are as accurate as those made by human experts. While the three methods provide very good results, locally weighted regression provides slightly better accuracy than the other methods. We also introduced a fourth method, a heterogeneous ensemble that combines the results of the three individual methods by means of an unweighted average. This ensemble performs better than any of the individual methods, both with the original data and when a signi£cant amount of Gaussian noise is added to the data.

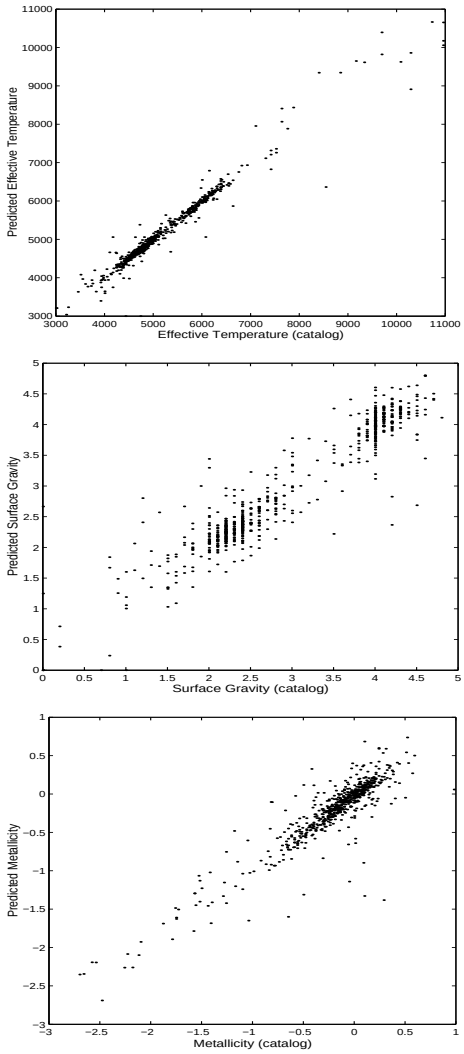

Figure 1. Catalog versus predicted parameters using LWR with the original dataset.

Future work includes extending the experiments to other datasets to test the generality of the conclusions reached here and work on developing methods for £nding the spectral indices that are the most useful for predicting each of the stellar atmospheric parameters.

References

K. Ali and M. Pazzani. Bagging predictors. *Machine Learning*, 24(3):173–202, 1996.

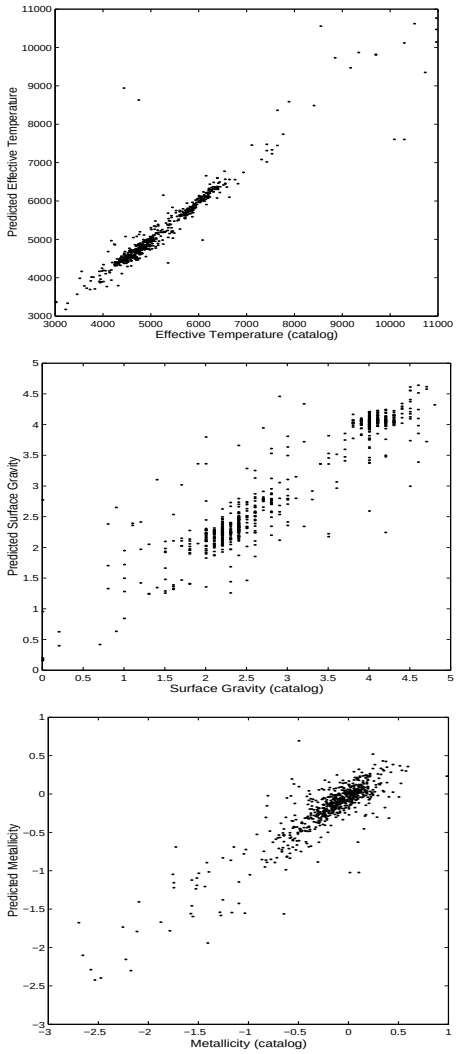

Figure 2. Catalog versus predicted parameters using a SOM with the original dataset.

C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Arti£cial Intelligence Review*, 11:11–73, 1997.

C. A. L. Bailer-Jones, M. Irwin, and T. von Hippel. Physical parameterization of stellar spectra: The neural network approach. *Monthly Notices of the Royal Astronomical Society*, 292(157), 1997.

C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, England, 1996.

L. Bouttou and V. Vapnik. Local learning algorithms. *Neural Computation*, 4(6):888–900, 1992.

L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

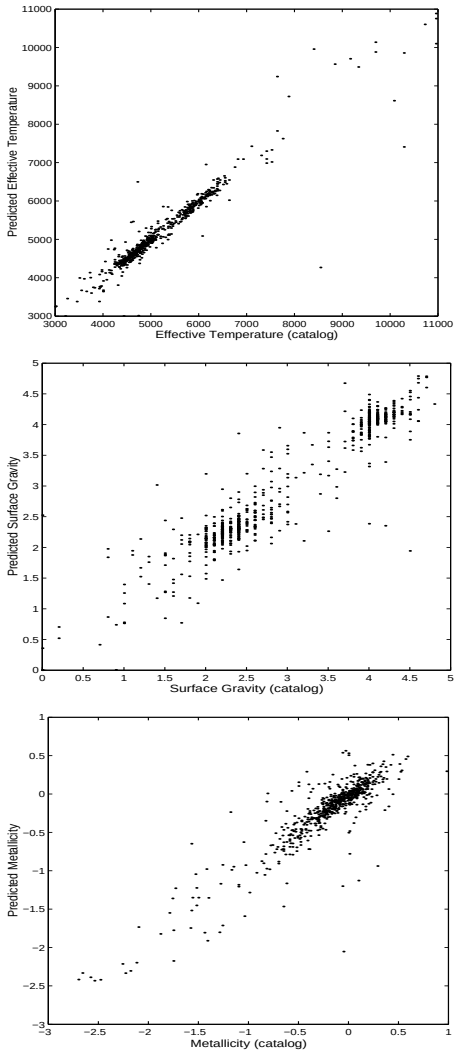

Figure 3. Catalog versus predicted parameters using with a FFNN with the original dataset.

A. J. Connolly, A. S. Szalay, M. A. Bershady, A. L. Kinney, and D. Calzetti. Spectral classi£cation of galaxies: an orthogonal approach. *The Astronomical Journal*, 110(3):1071–1082, 1995.

T. Dietterich and G.Bakiri. Solving multi-class learning problems via error-correcting output codes. *Journal of Arti£cial Intelligence Research*, 2:263–286, 1995.

S. R. Folkes, O. Lahav, and S. J. Maddox. An arti£cial neural network approach to classi£cation of galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, 283:651–665, 1996.

Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 148–156. Morgan Kaufmann, 1996.

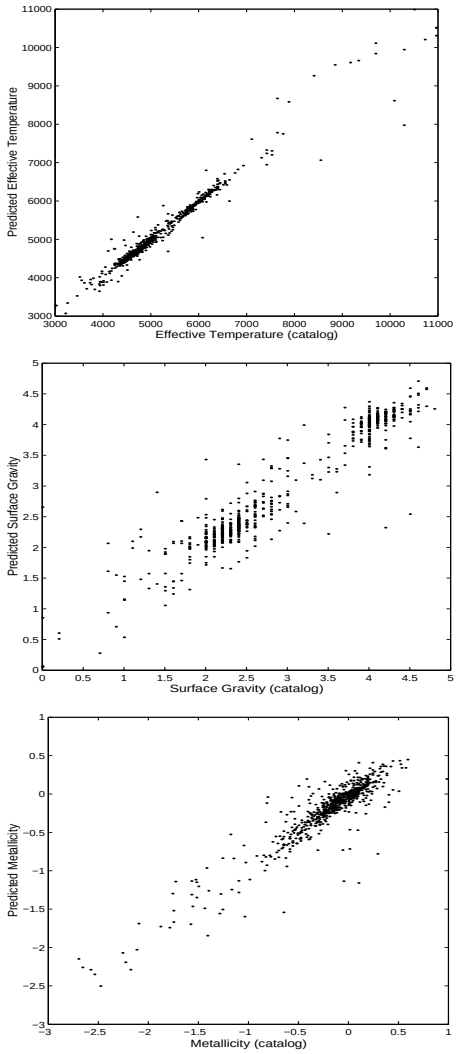

Figure 4. Catalog versus predicted parameters using the heterogeneous ensemble.

R. K. Gulati, R. Gupta, P. Gothoskar, and S. Khobragade. Stellar spectral classi£cation using automated schemes. *The Astrophysical Journal*, 426:340–344, 1994.

R. Hecht-Nielsen. *Neurocomputing*. Addison-Wesley, Reading, MA, 1990.

L. A. Jones. PhD thesis, University of North Carolina, Chapel Hill, North Carolina, 1996.

T. Kohonen. *Self-Organization and Associative Memory*. Springer-Verlag, Berlin, 1988. 2nd Edition.

S. W. Kwok and C. Carter. Multiple decision trees. In R. Schachter, T. S. Levitt, L. N. Kamal, and J. F. Lemmer, editors, *Uncertainty in Arti£cial Intelligence 4*, pages 327–335. Elsevier Science, Amsterdam, 1990.

S. C. Odewahn and M. L. Nielsen. Star-galaxy separation using neural networks. *Vistas in Astronomy*, 38:281–285, 1994.

M. C. Storrie-Lombardi, M. J. Irwin, T. von Hippel, and L. J. Storrie-Lombardi. Spectral classi£cation with principal component analysis and arti£cial neural networks. *Vistas in Astronomy*, 38:331–340, 1994.

M. C. Storrie-Lombardi, O. Lahav, L. Sodre, and L. J. Storrie-Lombardi. Morphological classi£cation of galaxies by arti£cial neural networks. *Monthly Notices of the Royal Astronomical Society*, 259:8–12, 1992.

W. B. Weaver and A. V. Torres-Dodgen. Neural network classi£cation of the near-infrered spectra of A-type stars. *The Astrophysical Journal*, 446:300–317, 1995.