Mon. Not. R. Astron. Soc. (1992) **259**, Short Communication, 8P–12P

Morphological classification of galaxies by Artificial Neural Networks

M. C. Storrie-Lombardi,[1] O. Lahav,[1] L. Sodré, Jr[2,3] and L. J. Storrie-Lombardi[1]

[1] *Institute of Astronomy, Madingley Road, Cambridge CB3 0HA*
[2] *Royal Greenwich Observatory, Madingley Road, Cambridge CB3 0EZ*
[3] *Instituto Astronômico e Geofísico da Universidade de São Paulo, CP9638, 01065, São Paulo, Brazil*

Accepted 1992 September 2. Received 1992 August 24

ABSTRACT

We explore a method for automatic morphological classification of galaxies by an Artificial Neural Network algorithm. The method is illustrated using 13 galaxy parameters measured by machine (ESO–LV), and classified into five types (E, S0, Sa + Sb, Sc + Sd and Irr). A simple Backpropagation algorithm allows us to train a network on a subset of the catalogue according to human classification, and then to predict, using the measured parameters, the classification for the rest of the catalogue. We show that the neural network behaves in our problem as a Bayesian classifier, i.e. it assigns the a posteriori probability for each of the five classes considered. The network highest probability choice agrees with the catalogue classification for 64 per cent of the galaxies. If either the first or the second highest probability choice of the network is considered, the success rate is 90 per cent. The technique allows uniform and more objective classification of very large extragalactic data sets.

Key words: methods: data analysis – catalogues – galaxies: fundamental parameters.

1 INTRODUCTION

The origin of the Hubble sequence remains a fundamental problem in understanding galaxy formation and the large-scale structure of the Universe. The morphological type describes the global appearance of a galaxy and provides useful information about its physical structure and the history of its stellar populations.

In spite of several attempts (e.g. Thonnat 1989; Okamura, Watanabe & Kodaira 1989; Lauberts & Valentijn 1989; Doi et al. 1992; Spiekermann 1992) to classify galaxies by deterministic algorithms, morphological classification into ellipticals, lenticulars, spirals and irregulars remains a process dependent on the eyes of a handful of dedicated individuals. We have investigated a computing technique, Artificial Neural Networks (ANNs), to classify galaxies. ANNs have several practical advantages. An ANN can be trained according to a subset classified by a human expert, and then it can classify the full data set. When more than one investigator contributes to the initial classification, the ANN learns each decision pattern, and produces a more uniform classification process free of such systematic errors as time effects. A uniform classification is also useful for producing target lists for surveys of selected type (e.g. a list of ellipticals for spectroscopic measurements of stellar velocity dispersion), and studies of morphological segregation on the large scale. Such automated procedures are the only practical way of classifying the enormous amount of data produced by machine scans of Schmidt plates like those obtained in the APM survey (Maddox et al. 1990).

Important by-products of developing an automated system include determining the primary physical parameters defining the Hubble sequence (cf. Brosche 1973; Meisels & Ostriker 1984), identifying new galaxy classes, and preserving human experience for a time-scale longer than the lifetime of a human expert.

2 ARTIFICIAL NEURAL NETWORKS

ANN algorithms, originally derived from simplified models of human central nervous system activity (McCullogh & Pitts 1943; Hopfield & Tank 1986), have found utility in astronomy for the classification of objects in the *IRAS* Point Source Catalog (e.g. Adorf & Meurs 1988), adaptive optics (e.g. Angel et al. 1990), scheduling observation time (e.g. Adorf 1989), and star–galaxy separation (e.g. Odewahn et al. 1991). Non-astronomical applications somewhat similar to our problem are speech recognition and identification of hand-written characters. For a review of these and other applications see, e.g., Gorman & Sejnowski (1988).

Here we use an ANN model known as the Backpropagation algorithm. It consists of nodes (analogous to human neurons) arranged in a series of layers. The nodes in a given layer are fully connected to the nodes in the next layer (see

Fig. 1). The input layer consists of the input parameters (13 in our case), and the output layer consists of the classes (five in our case). Any layer between the input and the output layers is called a 'hidden layer'. The input vector for each galaxy, containing the galaxy parameters, is presented to the network and the output is computed. The galaxy is then classified according to the class associated with the largest output component. The ANN can be viewed as a non-linear operator which transforms the distribution of objects in the input parameter space to the classification 'eigen-galaxies' space. The complexity (and non-linearity) of the ANN depends on the number of inputs, hidden nodes, layers, outputs and connections.

The network operates as follows. Each node (except the input nodes) receives the output of all nodes in the previous layer and produces its own output which then feeds the nodes in the next layer. A node at layer s calculates a linear combination over the input $x_i^{(s-1)}$ from the previous layer $s=1$ according to $I_j^{(s)}=\Sigma_i w_{ij}^{(s)} x_i^{(s-1)}$, where the w_{ij}'s are the weights associated with that node. The node then fires a signal $x_j^{(s)}=f(z)$ according to a non-linear threshold function usually of the sigmoid form $f(z)=1/[1+\exp(-z)]$ (in the interval [0, 1]) or $f(z)=\tanh(z)$ (in the interval $[-1, 1]$), where $z=I_j^{(s)}$.

For a given network architecture the first step is the 'training' of the ANN. In this step the weights w_{ij} (the 'free parameters') are determined by minimizing 'least-squares'. The novel aspect of Backpropagation is the way this minimization is done, using the chain rule (gradient descent) as proposed independently by several authors (e.g. Werbos 1974; Parker 1985; Rumelhart, Hinton & Williams 1986).

For each galaxy in the training set, the network compares its output vector in the 'classification space' $\boldsymbol{o}$ to the desired vector $\boldsymbol{d}$ determined by the human expert. The elements of the vector $\boldsymbol{d}$ are zero except for one element set to 1 corresponding to the actual class of the galaxy, e.g. we define $\boldsymbol{d}=(1, 0, 0, 0, 0)$ for ellipticals.

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

$$E=\frac{1}{2}\sum_k (o_k-d_k)^2,$$

where the sum is over the components of the vectors. This cost function, averaged over all the training galaxies presented to the ANN, is minimized with respect to free parameters, the weights w_{ij}. The weights are updated *backwards* from the output layer to one or more hidden layers by a small change in each time-step,

$$\Delta w_{ij}(t+1)=-\eta\frac{\partial E}{\partial w_{ij}}+\alpha\Delta w_{ij}(t),$$

where the 'learning coefficient' η and the 'momentum' α are 'knobs' which control the rate of learning of the network (see e.g. Hertz et al. 1991).

After completion of the 'learning' process by the use of a training set (i.e. fixing the weights w_{ij}) the ANN is ready to

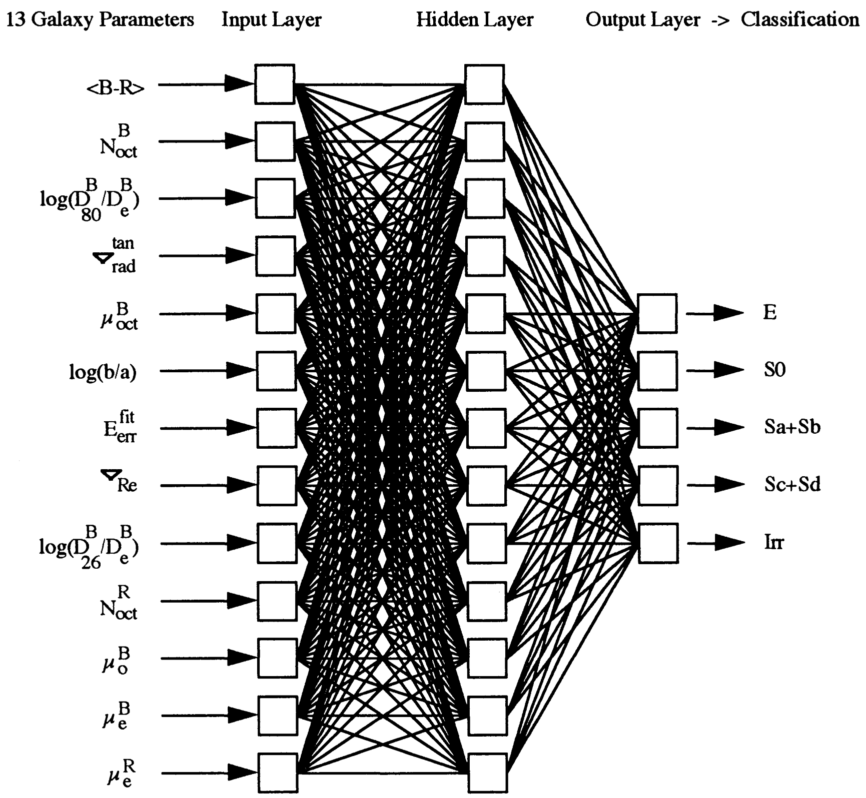

Figure 1. The ANN configuration (13; 13, 5) used in our study, with an input (galaxy parameters) layer of 13 nodes, a hidden layer of 13 nodes, and an output (classification) layer of 5 nodes. All nodes in a given layer are fully connected to all nodes in the next layer. The input parameters are explained in Table 1.

handle new unclassified data for which only the machine parameters are available. It then produces an output vector for each galaxy. The jth component of this vector can be viewed as the probability for class j given the input parameters $P(C_j|\boldsymbol{x})$. In fact, it can be proved theoretically (e.g. Gish 1990; Richard & Lippmann 1991) that the output of an ideal ANN is indeed a Bayesian a posteriori probability. Moreover, as our experiments confirm, the sum of the output vector components is $\sum_k o_k \approx 1$, as expected for a probabilistic classifier. It is worth noting that, unlike discrete classification of hand-written characters, galaxies form a continuous sequence. Hence the combination of probabilities assigned to different 'eigen-classes' may reflect an intermediate class.

We wish to emphasize that *supervised* ANNs do not produce an 'objective' unique classification. Supervised networks replicate the choices of their trainer – a network trained according to the classification made by Hubble or de Vaucouleurs will classify new data in a manner similar to the original expert.

3 EXPERIMENTS WITH ANN

Here we illustrate the method using the ESO–LV catalogue (Lauberts & Valentijn 1989, hereafter LV89). We have selected galaxies with ESO visual diameter ≥ 1 arcmin and at high Galactic latitude ($|b| > 30°$). Only galaxies with morphological classification performed by visual examination of the galaxy image are considered in our analysis. We use the 13 catalogue parameters shown in Table 1 to describe each galaxy. Hence, instead of going from 'pixels to galaxies', we have chosen to work with the much more compact information already contained in the ESO–LV catalogue. These 13 parameters were chosen because they are distance-independent, and they are very similar to those used by LV89 to perform the automated classification presented in the ESO–LV catalogue (hereafter ESO AUTO). This allows us to compare meaningfully the success rate of the classifications provided by our ANN with ESO AUTO. After selecting only galaxies with all 13 parameters available, our final data set has 5217 galaxies. We then randomly sort these galaxies in two independent sets of 1700 and 3517 objects for training and testing (samples TRAIN and TEST, respectively). We have also normalized our input data between 0 and 1 by using the minimum and maximum values of each parameter.

Table 1. The galaxy parameters.

- $\langle B-R \rangle$: average colour in region with B surface brightness 20.5 to 26;
- $N^B_{\rm oct}$: exponent of the fit of a generalized de Vaucouleurs law to B octants ($N=0.25$ corresponds to a perfect elliptical galaxy and $N=1$ to a pure exponential disc);
- $\log(D^B_{80}/D^B_{\rm e})$, where D^B_{80} and $D^B_{\rm e}$ are the major diameters of the ellipses at 80 per cent and half total B light, respectively;
- $\nabla^{\rm tan}_{\rm rad}$: arctangent of the absolute value of the ratio of the mean tangential and radial gradients, which is an indicator of the degree of asymmetry of the galaxy image;
- $\mu^B_{\rm oct}$: B central surface brightness from the fit of a generalized de Vaucouleurs law to B octants;
- $\log(b/a)$, where b/a is the galaxy axial ratio;
- $E^{\rm fit}_{\rm err}$: error in ellipse fit to B isophotes at B surface brightness 23;
- $\nabla_{\rm R_e}$: gradient of the B surface brightness profile at $D^B_{\rm e}$;
- $\log(D^B_{26}/D^B_{\rm e})$, where D^B_{26} is the major diameter of the ellipse at 26 B mag arcsec^{-2};
- $N^R_{\rm oct}$: exponent of the fit of a generalized de Vaucouleurs law to R octants;
- μ^B_0: average B surface brightness within 10 arcsec diameter circular aperture;
- $\mu^B_{\rm e}$: B surface brightness at half total B light;
- $\mu^R_{\rm e}$: R surface brightness at half total R light.

We have grouped the ESO–LV catalogue subclasses into five major classes and assigned each of the five output nodes of our networks to one of five classes of galaxies: E, S0, Sa+Sb, Sc+Sd, Irr. The distribution of the full set of 5217 galaxies as determined by LV89 is: E ($-5.0 \le T < -2.5$; 466 galaxies); S0 ($-2.5 \le T < 0.5$; 851 galaxies); Sa+Sb ($0.5 \le T < 4.5$; 2403 galaxies); Sc+Sd ($4.5 \le T < 8.5$; 1132 galaxies); and Irr ($8.5 \le T \le 10.0$; 365 galaxies), where T is the coded type.

We have investigated a variety of multilayer Backpropagation algorithms. All networks had 13 input nodes, one for each galaxy parameter, and five output nodes for classification. We present here results obtained with a very simple network, with only one hidden layer with 13 nodes. This configuration, labelled hereafter (13; 13, 5), is depicted in Fig. 1. We have used the sigmoid as our non-linear transfer function. The learning and momentum coefficients were kept constant at $\eta = 0.5$ and $\alpha = 0.2$, for all layers. We have verified, however, that our results are robust over a large range of these parameters. During training (using sample TRAIN), the ANN compared the output of these five nodes to the visual classification decisions of LV89. We then tested the network against the TEST sample. Morphological classification was performed by assigning the galaxy to the class corresponding to the maximal output component. It is worth mentioning that a classification scheme where one calculates the Euclidean distance of the ANN output from the vector representing each of the five possible classes, and then assigns the galaxy to the class producing the minimum vector distance, has produced exactly the same results (cf. Richard & Lippmann 1991).

Our main results, after 1 500 000 training iterations, are shown in Table 2, where we compare the visual and automated classifications for the TEST sample. Rows in the tables represent the visual type distribution, while the columns depict the automated type distribution. The diagonal presents the numbers of galaxies in each class for which human and automated procedures perfectly agree. From these tables one can verify that the percentage of galaxies correctly classified by ESO AUTO is 56 per cent. Our ANN, on the other hand, performs better: 64 per cent of the galaxies in the TEST sample were correctly classified. If

Table 2. Galaxy classification.

	(a) ANN					(b) ESO AUTO				
Class	E	S0	Sa+Sb	Sc+Sd	Irr	E	S0	Sa+Sb	Sc+Sd	Irr
E	203	77	25	1	5	197	87	17	5	5
S0	109	229	240	7	2	184	218	155	28	2
Sa+Sb	12	85	1281	218	15	106	12	791	664	38
Sc+Sd	1	4	304	415	36	22	11	24	631	72
Irr	0	0	53	69	126	22	9	31	42	144

we consider classification to within the nearest neighbour the success rate is much higher, 96 per cent.

Fig. 2 compares the galaxy morphological distribution produced by our network and by ESO AUTO with the visual distribution. As expected, both ESO AUTO and our network have difficulty agreeing with Lauberts & Valentijn about the decision boundaries between E and S0 and between Sc + Sd and Irr. However, ESO AUTO deviates dramatically from the visual perception of Lauberts & Valentijn in the distribution of Sa + Sb versus Sc + Sd, with ESO AUTO actually reversing the human finding of the number of Sa + Sb being larger than that of Sc + Sd in this sample. On the other hand, our network reproduces very well Lauberts & Valentijn's distribution of visual morphological types.

As discussed earlier, it can be shown that the ANN behaves like a Bayesian classifier (see e.g. Gish 1990; Richard & Lippmann 1991). Now we show that our network produces outputs which are indeed consistent with Bayesian a posteriori probabilities. First, in order to estimate probabilities, the network outputs should sum to 1 for each galaxy presented. Indeed, we find for the TEST sample that on the average $\sum_k o_k = 1.01 \pm 0.15$. Secondly, the output o_k *averaged over all inputs* should be the a priori class probability $P(C_k)$ for a class C_k. These expected values can be estimated by averaging the network outputs over all input data. Table 3 (for the TEST sample) compares the observed and estimated a priori class probabilities and indicates that our network correctly estimates these probabilities.

The probabilistic nature of the network not only provides insight into how the network operates, but also provides useful information on the classification quality of each individual galaxy. The distribution of $P_{\rm max}$, the value of the maximal output component, is different for galaxies correctly and wrongly classified. For the TEST sample, galaxies correctly classified have a median $P_{\rm max} \approx 0.84$, while galaxies wrongly classified have a median $P_{\rm max} \approx 0.71$, i.e. the ANN 'admits' making a fuzzier classification in this case. If either the first or the second highest outputs are considered in the comparison with the visual classification, the success rate is

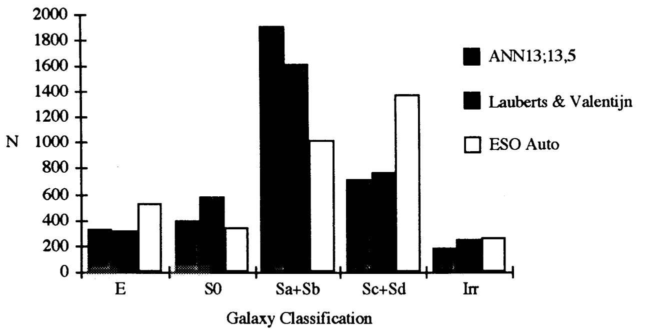

Figure 2. The classification of the TEST sample (3517 galaxies) according to the human eye (LV), ESO AUTO, and our ANN (13; 13, 5). ESO AUTO exhibits Sc + Sd excess and underestimates Sa + Sb as compared to human and ANN classifications.

Table 3. The a priori class probability.

Class	E	S0	Sa+Sb	Sc+Sd	Irr
L&V Visual	0.088	0.167	0.458	0.216	0.071
ANN	0.093	0.152	0.482	0.216	0.062

90 per cent. We have also found that filtering of ill-defined galaxies in the training set further improves the classification.

4 DISCUSSION

We have illustrated that the ANN artificial intelligence method is able to produce useful galaxy classification by assigning Bayesian probabilities to each possible morphological type. In spite of the facts that ESO–LV is based on plate material, the training set was produced by several observers, and the galaxy parameters were chosen somewhat arbitrarily, the ANN predicted reasonably well the morphological type of galaxies. Clearly, by using CCD frames, a uniform training set, and a more optimal set of galaxy parameters, one can improve the classification further.

The ANN classification method improves considerably on statistical techniques commonly used. ANN algorithms make no prior assumptions about the statistical distribution of test objects, and invoke no heuristics to help define class membership. Our initial success with simple Backpropagation has encouraged us to pursue other aspects of ANN technique, including: (i) determining the optimal choice of network parameters, e.g. the number of hidden layers and nodes, and learning and momentum coefficients; (ii) assessing the contribution of the galaxy characteristics, and finding the 'best parameters'; (iii) finding which fundamental parameters are defining the Hubble sequence; (iv) utilizing the Automated Plate Measuring (APM) facility in Cambridge to provide ANNs with the input of full 2D pixel maps of thousands of galaxies, and (v) producing catalogues of galaxies with assigned Bayesian probability of morphological classification.

ACKNOWLEDGMENTS

We thank W. Fitzgerald, J. Hertz, M. Irwin and D. Lynden-Bell for helpful discussions. LSJ thanks FAPESP and USP/BID and MCSL thanks the Sheepshanks Fund for financial support.

REFERENCES

Adorf H. M., 1989, in Heck A., Murtagh F., eds, Knowledge-Based Systems in Astronomy. Springer-Verlag, Heidelberg, p. 215

Adorf H. M., Meurs E. J. A., 1988, in Seitter W. C., Duerbeck H. W., Tacke M., eds, Large-Scale Structure of the Universe: Observational and Analytical Methods. Springer-Verlag, Heidelberg, p. 315

Angel J. R. P., Wizinowich P., Lloyd-Hart M., Sandler D., 1990, Nat, 348, 221

Brosche P., 1973, A&A, 23, 234

Doi M., Kashikawa N., Okamura S., Tarusawa K., Fukugita M., Sekiguchi M., Iwashita H., 1992, in MacGillivray H. T., Thomson E. B., eds, The 2nd Meeting on Digitized Optical Sky Survey. Kluwer, Dordrecht

Gish H., 1990, in Proc. IEEE Conf. on Acoustics Speech and Signal Processing, p. 1361

Gorman R. P., Sejnowski T. J., 1988, Neural Networks, 1, 75

Hertz J., Krogh A., Palmer R. G., 1991, Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, California

Hopfield J. J., Tank D. W., 1986, Sci, 233, 625

Lauberts A., Valentijn E. A., 1989, The Surface Photometry Catalogue of the ESO–Uppsala Galaxies. ESO (LV89)

Maddox S. J., Sutherland W. J., Efstathiou G., Loveday J., 1990, MNRAS, 243, 692
McCullogh W. S., Pitts W. H., 1943, Bull. Math. Biophys., 5, 115
Meisels A., Ostriker J. P., 1984, AJ, 89, 1451
Odewahn S. C., Stockwell E. B., Pennington R. L., Humphreys R. M., Zumach W. A., 1991, AJ, 103, 318
Okamura S., Watanabe M., Kodaira K., 1989, in Corwin H. G., Jr, Bottinelli L., eds, The World of Galaxies. Springer-Verlag, New York, p. 75
Parker D. B., 1985, Report TR-47. MIT Center for Computational Research in Economics and Management Science, Cambridge, MA
Richard M. D., Lippmann R. P., 1991, Neural Comput., 3, 461
Rumelhart D. E., Hinton G. E., Williams R. J., 1986, Nat, 323, 533
Spiekermann G., 1992, AJ, 103, 2102
Thonnat M., 1989, in Corwin H. G., Jr, Bottinelli L., eds, The World of Galaxies. Springer-Verlag, New York, p. 53
Werbos P. J., 1974, PhD thesis, Harvard University