Mon. Not. R. Astron. Soc. **275**, 567–590 (1995)

Automated morphological classification of APM galaxies by supervised artificial neural networks

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

[1] *Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA*
[2] *Instituto Astronômico e Geofísico da Universidade de São Paulo, CP9638, 01065-970, São Paulo, Brazil*

Accepted 1995 January 31. Received 1995 January 27; in original form 1994 November 3

ABSTRACT

We train artificial neural networks to classify galaxies based solely on the morphology of the galaxy images as they appear on blue survey plates. The images are reduced, and morphological features such as bulge size and the number of arms are extracted, all in a fully automated manner. The galaxy sample was first classified by six independent experts. We use several definitions for the mean type of each galaxy, based on those classifications. We then train and test the network on these features. We find that the rms error of the network classifications, as compared with the mean types of the expert classifications, is 1.8 Revised Hubble types. This is comparable to the overall rms dispersion between the experts. This result is robust and almost completely independent of the network architecture used.

Key words: catalogues – galaxies: general.

1 INTRODUCTION

Since the introduction of the Hubble classification scheme (Hubble 1926, 1936), astronomers have been looking at ways to classify galaxies. Other systems were suggested, e.g. Mt. Wilson (Sandage 1961), Yerkes (Morgan 1958), Revised Hubble (de Vaucouleurs 1959) and DDO (van den Bergh 1960a,b, 1976); each has its special characteristics, but they all share Hubble's original notion that the sequence of morphologies attests to an underlying sequence of physical processes.

This notion has been widely accepted for the past few decades, making morphological classification of large numbers of galaxies important for better modelling and understanding of galaxy structure and evolution. Examples include statistical relations which are specific to certain types of galaxies, e.g. the D_n–σ relation for ellipticals (Lynden-Bell et al. 1988), the Tully–Fisher relation for spirals (Tully & Fisher 1977) and the morphology–density relation (Hubble 1936; Dressler 1980).

Morphological classification of galaxies is usually done by visual inspection of photographic plates. This is by no means an easy task, requiring skill and experience. It is also time-consuming: catalogues containing human classifications take years to complete and contain of order 10^4 entries [e.g. the Third Reference Catalogue of Bright Galaxies (de Vaucouleurs et al. 1991) and the ESO catalogue (Lauberts & Valentijn 1989)]. However, in the APM (Automated Plate Measuring machine) survey (e.g. Maddox et al. 1990) there are roughly 2×10^6 galaxies, and the expected yield of the Sloan Digital Sky Survey (Gunn et al., in preparation) is over 10^7 CCD images of galaxies. Clearly, such numbers of galaxies cannot be classified by humans. There is an obvious need for automated methods that will put the knowledge and experience of the human experts to use and produce very large samples of automatically classified galaxies.

The first stage towards achieving this goal was creating a uniform, well-defined sample to be classified by human experts. This was done in previous papers (Lahav et al. 1994; Naim et al. 1995, hereafter Paper I), where the same sample of galaxies was presented to six independent expert observers and a detailed analysis of their classifications was carried out. The experts are R. Buta, H. Corwin, G. de Vaucouleurs, A. Dressler, J. Huchra and S. van den Bergh (hereafter RB, HC, GV, AD, JH and vdB, respectively). We found that the rms dispersions between pairs of experts ranged from 1.3 to 2.1 Revised Hubble types, and that the overall rms dispersion was 1.8 types.

The next stage, which is carried out in this paper, entails training a computer software to classify galaxies on the basis of their apparent morphology. Our choice of an automated classifier is artificial neural networks (ANNs), which proved in a pilot study (Storrie-Lombardi et al. 1992) to be well suited for this task. The original idea behind ANNs was the creation of a simplified model of the human brain (McCulloch & Pitts 1943; Hopfield & Tank 1986), but they