Summary

“Using Oblique Decision Trees for the Morphological Classification of Galaxies”
(E.A. Owens, R.E. Griffiths, and K.U. Ratnatunga)

Owens, Griffiths, & Ratnatunga (1996) used the combined work of other researchers in order to formulate a "comparable" algorithm for the morphological classification of galaxies. They implemented an oblique decision tree algorithm formulated by Murthy, Kasif, & Salzberg (1994) originally shown to produce accurate decision trees for the purpose of distinguishing between galaxies and stars or identifying cosmic rays in telescope images. This is applied to galactic morphology data provided by Storrie-Lombardi et al. (1992) (referred to as SLSS), which was originally taken from the ESO-LV galaxy catalog [1], and compared to results from a neural network classifier created by SLSS. Before discussing the specific methods used, Owens et al. point out the fundamental differences of the dominant aspects of the implemented neural networks and decision trees. They determine that it is more accurate to group neighboring classes and suggest some possibly significant sources of error.

The SLSS ANN (artificial neural network) has 13 input nodes (corresponding to various parameters determined from images of galaxies), 13 hidden nodes, and 5 output nodes. The 5 outputs correspond to five specific classes chosen by SLSS: E, S0, Sa+Sb, Sc+Sd, and Irr [2]. Error minimization is done through back-propagation, or modifying the weights of nodes towards lower error values. Final classification is given by the node with the largest output value. The data set of 5217 images was randomly divided into a training set (1700 images) and a test set (3517 images).

The decision trees constructed in this experiment are called oblique decision trees. The data space is divided in a binary fashion (i.e. ‘yes’ or ‘no’) thus positively or negatively classifying objects; in this case, the classification is done by comparison to a threshold value. Tests of this form are called splits. Splits in which only one parameter is examined (segmentation along one dimension) are called axis-parallel, while those which compare a linear combination of two or more parameters (segmentation along a multi-dimensional hyperplane) are called oblique [3]; consequently, a decision tree with one or more oblique splits is considered oblique. Owens et al. use an extension of Murthy et al.’s Oblique Classifier 1 (OC1) algorithm for the tree construction. OC1 forms decision trees by determining the quality of each split given by a measure of impurity, or (heuristically) how “poorly” data is separated. Since finding an optimal oblique split is an NP-Complete problem, a stochastic greedy search is done instead (random, best-match, with no look-ahead). The random element is introduced in order to prevent the possibility of analyzing data that is too “localized” to one part of the large data set provided, and it is also mentioned that randomization helps produce a “most popular” result through the generation of multiple decision trees (a “decision forest”). Owens et al. used the same training and test data as was used in the SLSS ANN.

Running OC1 in axis-parallel mode resulted is a 1% difference from the SLSS results. The most accurate of five random tests (not in axis-parallel mode) produced a smaller tree with the same accuracy (but a lower average). The authors noted that the resulting tree is still comparatively large, and that they managed to formulate a tree composed of only 6 leaves (compared to an average of 17 for the five random runs) using a different information measure. While they expected OC1’s smaller trees to be more accurate (due to Ocam’s razor), the larger trees were in fact the more accurate – this phenomenon was tentatively correlated to the number of leaves, indicating data that is complex and difficult to generalize.
The results indicated, as in the SLSS ANN, that non-neighboring classes are more accurately separated than neighboring classes. They explain how using different combinations of multiple classifications increases accuracy, especially for the E, Sa+Sb, and Irr grouping. Overall errors are attributed to either visual classification errors (in the catalog), or insufficient information provided by the chosen features (which were apparently chosen “somewhat arbitrarily”). They finally comment that the correlation of mis-classification between the two algorithms supports the possibility of error in the original classification and confirms the SLSS result that the distinction between neighboring classes is poorly defined.

References.


