

Photometric identification of quasars from the Sloan Survey

Rameshwar P. Sinha¹, Ninan S. Philip², Ajit K. Kembahvi¹ and Ashish A. Mahabal³

¹Inter-University Centre for Astronomy and Astrophysics (IUCAA),
Post Bag 4, Ganeshkhind, Pune University Campus, Pune 411 007, India
email: rsinha,akk@iucaa.ernet.in

²St. Thomas College, Kozhencheri, Kerala, India

³Astronomy Department, California Institute of Technology, Pasadena, CA 91125, USA
email: milan@astro.caltech.edu

We have developed a neural network based technique for identifying quasars from the photometric database of the Sloan Digital Sky Survey (SDSS). We have queried the SDSS data release 5 (DR5) to produce a dataset of spectroscopically identified samples of unresolved objects consisting of quasars and stars which forms the training set.

Using a representative subsample of this dataset we train a Difference Boosting Neural Network (DBNN) that looks for differences in the colours of the objects to classify them. The four colours, namely $u-g$, $g-r$, $r-i$, $i-z$ and the spectral classes, i.e., star and quasar are used as the inputs for the network which returns a predicted spectral class of each object, assigning a confidence level to the prediction.

We use a few subsamples of 10 000 randomly selected objects from the full dataset and train the network. Subsequently, we divide the whole colour space into different sections each having a different density of quasars and again train the DBNN. After a few rounds of training with each subsample set we get the best trained neural net based on adaptive data selection.

The quality of training data is adjudged only on the basis of the accuracy achieved when the trained data is tested on unseen data. We use our best trained network to find quasars from the test dataset which consists of spectroscopically confirmed quasars and stars. From the results we estimate the level of completeness and the degree of contamination to the quasars identified using the DBNN. We report our best trained network which has a completeness of 98% and a contamination of 2% from stars. The trained network can be used to identify quasars from samples of unresolved objects selected from DR5.

We have made use of VO tools such as VOPlot and VOPlot-3D. In particular, VOPlot-3D has helped in visually demonstrating the separation between stars and quasars in a multi-dimensional colour space as well as highlighting the hypersurface created by the DBNN for each of the two classes. We also make use of few statistical tools such as box-plot, histogram and k -mean clustering from VOSTat to substantiate our results further and facilitate our investigation. In the near future, we plan to make a web-based tool with a user-friendly interface for making selections using DBNN. The technique can have a wide application.