Co-Evolution of Central Black Holes and Galaxies
Proceedings IAU Symposium No. 267, 2009
B. M. Peterson, R. S. Somerville, & T. Storchi-Bergmann, eds.

doi:10.1017/S1743921310006083

Automated Classification of Quasars and Stars

Yanxia Zhang[1], Yongheng Zhao[1], and Hongwen Zheng[2]

[1]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
Email: zyx@lamost.org

[2]North China Electronic Power University, Beijing, 102206, China
Email: zyx300@sohu.com

Keywords. astronomical databases: miscellaneous, catalogs, methods: data analysis, methods: statistical

We investigate selection and weighting of features by applying a random forest algorithm to multiwavelength data. Then we employ a k-nearest neighbor method to distinguish quasars from stars. We then compare the performance of this approach based on all features, weighted features, and selected features. We find that the k-nearest neighbor approach combined with random forests effectively separates quasars from stars.

The sample we used was cross-identified from different survey catalogs, i.e., the SDSS DR5, FIRST, and USNO-B1.0 catalogs. This yielded as sample of 6,479 quasars and 785 stars.

A random forest approach was used to compute a weight for each attribute which allows us to select the most important attributes. We used the k-nearest neighbor approach to discriminate between quasars and stars and the results for three different variants are shown in Table 1. The accuracy for quasar selection is above 98% for all three variants, but the classification of stars is not as good. The overall accuracy is better than 89% and in the best case the total accuracy is 94.93%. Table 1 also shows that the performance with weighted features or selected features is slightly better than that with all features. As a consequence, if we have many input features, we generally need selection or weighting of features before we begin k-NN model building.

Table 1. Separate quasars and stars by k-NN.

Sample	All	Features	Weighted	Features	Selected	Features
classified↓known→	quasars	stars	quasars	stars	quasars	stars
quasars	6440	705	6373	345	6373	262
stars	39	80	106	440	106	523
Accuracy(%)	99.4±0.4	10.2±2.3	98.4±0.4	56.0±5.1	98.4±0.6	66.6±5.3
Total accuracy(%)	89.76±0.55		93.79±0.68		94.93±0.86	

Acknowledgements

This work has been funded by the National Natural Science Foundation of China under Grant Nos. 90412016 and 10778724, and by 863 project under Grant No. 2006AA01A120.