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/S1743921310006071

Discrimination Between Quasars and Stars by Support Vector Machines

Yanxia Zhang[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 investigated the use of support vector machines (SVMs) to distinguish quasars from stars using survey databases from different wavebands. We first employed a random forest approach for selection and weighting of features. We find that SVMs are an effective method to preselect quasar candidates from multiwavelength data.

Two data sets were used, one from SDSS, USNO-B1.0, and FIRST (the "FIRST sample"), and another from SDSS, USNO-B1.0, and *ROSAT* (the "ROSAT sample"). We analyzed SVM performance for different features in Tables 1 and 2 and find that the accuracy with the FIRST sample was superior to that with the ROSAT sample except in the case of feature selection. Performance was improved by using selected features, but was poorer when features were weighted. We conclude that when SVMs are used for classification purposes, feature selection is necessary since this not only improves the performance, but also reduces the dimensionalities.

Table 1. Separation of quasars and stars in the FIRST sample by SVMs.

Sample	All	Features	Weighted	Features	Selected	Features
classified↓known→	quasars	stars	quasars	stars	quasars	stars
quasars	6479	779	6479	781	6426	495
stars	0	6	0	4	53	290
Accuracy(%)	100.0±0.00	0.01±0.00	100.0±0.00	0.01±0.00	99.18±0.18	36.91±2.83
Total accuracy(%)	89.28±0.00		89.25±0.00		92.46±0.40	

Table 2. Separation of quasars and stars in the ROSAT sample by SVMs.

Sample	All	Features	Weighted	Features	Selected	Features
classified↓known→	quasars	stars	quasars	stars	quasars	stars
quasars	6347	1587	6360	2355	6221	71
stars	13	770	0	2	139	2286
Accuracy(%)	99.80±0.12	32.65±3.73	100.0±0.00	0.00±0.00	97.81±0.77	96.99±1.15
Total accuracy(%)	81.64±0.99		72.98±0.08		97.60±0.56	

Acknowledgements

This paper 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.