Astrophysics from Antarctica
Proceedings IAU Symposium No. 288, 2012
M. G. Burton, X. Cui & N. F. H. Tothill, eds.

doi:10.1017/S1743921312017176

Classification of Quasars and Stars by Supervised and Unsupervised Methods

Yanxia Zhang[1], Yongheng Zhao[1], Hongwen Zheng[2] and Xue-bing Wu[3]

[1]Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China
email: zyx@bao.ac.cn,yzhao@bao.ac.cn

[2]Mathematics and Physics Department, North China Electronic Power University, Beijing 102206, China

[3]Department of Astronomy, Peking University, Beijing 100871, China

Abstract. Targeting quasar candidates is always an important task for large spectroscopic sky survey projects. Astronomers never give up thinking out effective approaches to separate quasars from stars. The previous methods on this issue almost belong to supervised methods or color-color cut. In this work, we compare the performance of a supervised method – Support Vector Machine (SVM)– with that of an unsupervised method one-class SVM. The performance of SVM is better than that of one-class SVM. But one-class SVM is an unsupervised algorithm which is helpful to recognize rare or mysterious objects. Combining supervised methods with unsupervised methods is effective to improve the performance of a single classifier.

Keywords. Classification, Astronomical databases: miscellaneous, Catalogs, Methods: data analysis, Methods: statistical

1. Introduction

Large samples of quasars are important tools in astrophysics and cosmology for several reasons. With them, we may not only study the quasar phenomenon itself, but also the numerous astrophysical applications they offer. The quasar phenomenon itself is related to the galaxy history and evolution, to star formation and possibly to the interaction with other galaxies. Quasars can be used to study the intervening intergalactic medium. Indeed, the mechanism which feeds the central black hole and triggers the nucleus active during a given period remains unclear. So far there have been many automated methods focusing on selecting quasar candidates.

2. Sample and results

The samples applied here were cross-identified from different survey catalogs: SDSS DR7 and UKIDSS DR7 catalogs. The star sample was adopted from the cross-identified pointed sample without brightness variation in Stripe 82. Finally we obtain 21,241 quasars and 154,739 stars.

Support Vector Machine (SVM) is a two-class algorithm (i.e. one needs negative as well as positive examples). One-class SVM focuses on positive data, and identifies outliers amongst the positive examples and uses them as negative examples. SVM and one-class SVM are compared to separate quasars from stars. The input pattern for SVM and one-class SVM is $i, u-g, g-r, r-i, i-z, z-Y, Y-J, J-H, H-K$. The classification results are shown in Table 1. Obviously the performance of SVM is superior to that of one-class SVM. The accuracy of quasars and stars with SVM is more than 99.9%, respectively. The accuracy of quasars with one-class SVM is more than 99.8% while the

accuracy of stars is rather poor. Comparing the misclassified sources by the two methods, the misclassified quasars have no overlap. However misclassified stars by one-class SVM include misclassified stars by SVM. As a result, we propose a new classification scheme as indicated in Figure 1.

Table 1. The classification results with SVM and one-class SVM

method	SVM		one-class SVM	
classified↓known→	quasars	stars	quasars	stars
quasars	21,226	32	21,204	66,792
stars	15	154,707	37	87,948
accuracy	99.93%	99.98%	99.83%	56.80%

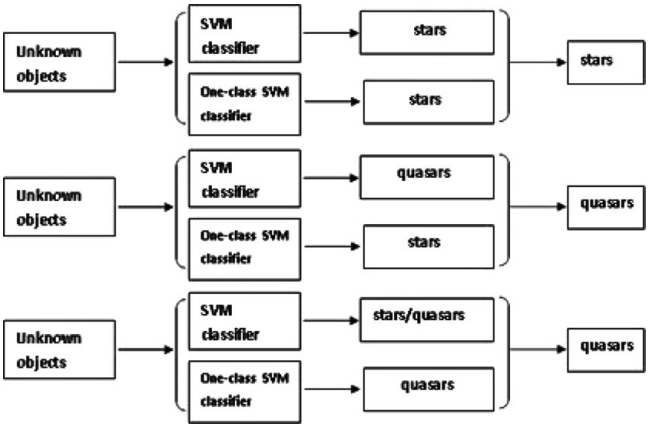

Figure 1. Classification scheme of combined SVM and one-class SVM.

Table 2. The classification result with combined SVM and one-class SVM

classified↓known→	quasars	stars
quasars	21,241	32
stars	0	154,707
accuracy	100.00%	99.98%

When combining the two approaches, the new objects are recognized as stars when SVM classifier and one-class SVM classifier both consider them as stars; the new objects are marked as quasars when SVM or one-class SVM classifies them as quasars. The new classification accuracy based on this scheme improves as shown in Table 2. In other words, ensemble different methods are helpful to enhance the performance of a single classifier.

Acknowledgements

This paper is funded by the National Natural Science Foundation of China under grant No.10778724, 11178021 and No. 11033001, the Natural Science Foundation of Education Department of Hebei Province under grant No. ZD2010127. We acknowledge use of the SDSS and UKIDSS databases.