Research in Astron. Astrophys. **2009** Vol. **9** No. **2**, 220–226
http://www.raa-journal.org *http://www.iop.org/journals/raa*

Research in Astronomy and Astrophysics

Random forest algorithm for classification of multiwavelength data *

Dan Gao[1,2], Yan-Xia Zhang[1] and Yong-Heng Zhao[1]

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

[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Received 2008 February 27; accepted 2008 May 12

Abstract We introduced a decision tree method called Random Forests for multiwavelength data classification. The data were adopted from different databases, including the Sloan Digital Sky Survey (SDSS) Data Release five, USNO, FIRST and ROSAT. We then studied the discrimination of quasars from stars and the classification of quasars, stars and galaxies with the sample from optical and radio bands and with that from optical and X-ray bands. Moreover, feature selection and feature weighting based on Random Forests were investigated. The performances based on different input patterns were compared. The experimental results show that the random forest method is an effective method for astronomical object classification and can be applied to other classification problems faced in astronomy. In addition, Random Forests will show its superiorities due to its own merits, e.g. classification, feature selection, feature weighting as well as outlier detection.

Key words: classification — astronomical databases: miscellaneous — catalogs — methods: data analysis — methods: statistical

1 INTRODUCTION

With the development and implementation of large space-based and ground-based survey projects, such as 2dF, 2MASS, NVSS, FIRST and SDSS, the amounts of astronomical data and information become larger and larger. How to extract knowledge from a huge volume of data by automated methods is an important task for astronomers. The automatic classification of objects from catalogs or other sources of data is a common statistical problem in many astronomical surveys. So far, there has been much work done on this issue. The Neural Network algorithm was used for spectral classification of galaxies (Sodré & Cuevas 1994) and for morphological classification of galaxies (Storrie-Lombardi et al. 1992; Adams & Woolley 1994). Learning Vector Quantization (LVQ) was applied to the classification of astronomical objects classification (Zhang & Zhao 2003). Bayesian Belief Networks (BBN), Multilayer Perceptron (MLP) networks and Alternating Decision Trees (ADtree) were compared for their ability to separate quasars from stars (Zhang & Zhao 2007). Support vector machines (SVMs) have also been successfully applied to automatic classification (Zhang & Zhao 2003, 2004). Decision trees, e.g. REPTree, Random Tree, Decision Stump, Random Forest, J48, NBTree and ADTree were investigated to classify active objects from non-active objects (Zhao & Zhang 2007).

In this paper, we explore an effective method, Random Forests, in which votes for class membership are polled from a large random ensemble of tree classifiers. The random forest method has many

* Supported by the National Natural Science Foundation of China.

successful applications in astronomy, for example, Breiman et al. (2003) used this method to identify quasars from the FIRST survey. Albert (2007) applied it in the analysis of data from the ground-based gamma telescope MAGIC. Carliles et al. (2007) employed it to estimate galaxy redshifts with color features and spectroscopic redshifts from SDSS DR6. Bailey et al. (2007) presented the classification results of Random Forests to different images in the context of the Nearby Supernova Factory supernova search. This paper is organized as follows: Section 2 introduces the principal of Random Forests; Section 3 describes the samples used; the results and discussions are given in Section 4; finally the conclusion summarizes this paper in Section 5.

2 METHOD

A random forest is an ensemble (i.e., a collection) of unpruned decision trees. Random forests are often used when we have very large training datasets and a very large number of input variables (hundreds or even thousands of input variables). A random forest model is a classifier that consists of many decision trees and outputs the class that is the mode of the class output by individual trees (Breiman 2001). Regarding the principle and software of Random Forests, we refer readers to the website (*http://www.stat.berkeley.edu/users/breiman/RandomForests/cc_home.htm*).

Each tree is constructed using the following algorithm:

1. Let the number of training cases be N, and the number of variables in the classifier be M.
2. We are told the number m of input variables to be used to determine the decision at a node of the tree; m should be much less than M.
3. Choose a training set for this tree by choosing N times with replacement from all N available training cases (i.e. take a bootstrap sample). Use the rest of the cases to estimate the error of the tree, by predicting their classes.
4. For each node in the tree, randomly choose m variables on which to base the decision at that node. Calculate the best split based on these m variables in the training set.
5. Each tree is fully grown and not pruned (as may be done in constructing a normal tree classifier).

The advantages of random forests are:

• For many data sets, it produces a highly accurate classifier.

• It handles a very large number of input variables.

• It estimates the importance of variables in determining classification.

• It generates an internal unbiased estimate of the generalization error as the forest building progresses.

• It includes a good method for estimating missing data and maintains accuracy when a large proportion of the data are missing.

• It provides an experimental way to detect variable interactions.

• It can balance error in the class population of unbalanced data sets.

• It computes proximities between cases, useful for clustering, detecting outliers, and (by scaling) visualizing the data.

• Using the above, it can be extended to unlabeled data, leading to unsupervised clustering, outlier detection and data views.

• Learning is fast.

3 SAMPLE AND PARAMETER SELECTION

The sample we used in this paper was cross-identified from different survey catalogs, i.e. the Sloan Digital Sky Survey (SDSS) DR5, the Faint Images of the Radio Sky at Twenty centimeters (FIRST), the ROSAT All-Sky survey (RASS) and USNO-B1.0 catalogs.

The SDSS survey (York et al. 2000) uses a dedicated, wide field, 2.5 m telescope at Apache Point Observatory, New Mexico. Imaging is carried out in drift-scan mode using a 142 mega-pixel camera in five broad bands, u, g, r, i and z, spanning the range from 3000 to 10 000 Å. The corresponding magnitude limits for the five bands are 22.0, 22.2, 22.2, 21.3 and 20.5, respectively. The Fifth Data Release (DR5) of the SDSS includes all survey quality data taken through June 2005 and represents the completion of the SDSS-I project. It includes five-band photometric data for 215 million unique objects selected over 8000 deg^2 , and 1 048 960 spectra of galaxies, quasars, and stars selected from 5740 deg^2 of that imaging data.

The FIRST survey began in 1993. It uses the VLA (Very Large Array, a facility of the National Radio Observatory (NRAO)) at a frequency of 1.4 GHz and it is slated to observe 10 000 deg^2 of the North and South Galactic Caps, to a sensitivity of about 1 mJy with an angular resolution of about 5 arcsec. The images produced by an automated mapping pipeline have pixels of 1.8 arcsec, a typical rms of 0.15 mJy, and a resolution of 5 arcsec; the images are available on the Internet (see the FIRST home page at *http://sundog.stsci.edu/* for details). The source catalog is derived from the images. A new catalog (Becker et al. 2003) of the FIRST Survey has been released that includes all data taken from 1993 through September 2002, and contains about 811 000 sources covering 8422 deg^2 in the North Galactic cap and 611 deg^2 in the South Galactic cap. The new catalog and images are accessible via the FIRST Search Engine and the FIRST Cutout Server.

The RASS survey uses an imaging X-ray Telescope (Trumper 1983) and the data are well suited for investigating the X-ray properties of astronomical objects. The RASS Bright Source Catalog (RBSC) includes 18 811 sources, with a limiting ROSAT PSPC count rate of 0.05 counts s^{-1} in the 0.1–2.4 keV energy band (Voges et al. 1999). The typical positional accuracy is 30 arcsec. Similarly, the RASS Faint Source Catalog (RFSC) contains 105 924 sources and represents the faint extension to the RBSC (Voges et al. 2000). The RBSC and RFSC catalogs contain the ROSAT name, position in equatorial coordinates, the positional error, the source count rate (CR) and error, the background count rate, exposure time, hardness-ratios (HR1 and HR2) and their errors, extent (ext) and likelihood of extent (extl) and likelihood of detection. From the count rate A in the 0.1–0.4 keV energy band and the count rate B in the 0.5–2.0 keV energy band, HR1 is given by: HR1 = (B-A)/(B+A). HR2 is determined from the count rate C in the 0.5–0.9 keV energy band and the count rate D in the 0.9–2.0 keV energy band by: HR2 = (D-C)/(D+C). CR is the ROSAT total count rate in counts s^{-1}. The parameters of the ext and extl are the source extent in arcseconds and the likelihood of the source extent in arcseconds, respectively. The ext is the amount by which the source image exceeds the point spread function. The parameters of the ext and extl reflect whether sources are point sources or extended sources. For example, stars or quasars are point sources; galaxies or galaxy clusters are extended sources. Thus the ext and extl are useful for classification of these kinds of objects.

The USNO-B1.0 is a catalog that presents positions, proper motions, magnitudes in various optical passbands and star/galaxy estimators for 1 045 913 669 objects derived from 3 648 832 040 separate observations (Monet et al. 2003). The data were taken from scans of 7 435 Schmidt plates taken from various sky surveys during the last 50 years. The catalog is expected to be complete down to $V = 21$; the estimated accuracies are 0.2 arcsec for the positions at J2000, 0.3 mag in up to 5 colors, and 85% accuracy for distinguishing stars from non-stellar objects.

After cross-matching, we collect 6479 quasars, 785 stars and 27 984 galaxies with SDSS DR5, FIRST and USNO-B1.0 measurement as the first sample (S1), and also get 6362 quasars, 2357 stars and 5333 galaxies with SDSS DR5, ROSAT and USNO-B1.0 measurement as the second sample (S2). The description of the samples is listed in Table 1.

We choose the five model magnitudes from SDSS DR5 and correct the magnitudes by Galaxy extinction using the dust maps of Schlegel et al. (1998). All the following SDSS magnitudes (u, g, r, i, z) or colors $(u-g, g-r, r-i, i-z)$ are dealt with in this way. The attributes from the FIRST survey are

Table 1 Description of Samples

Sample	Catalogs	No. of Quasars	No. of Stars	No. of Galaxies
S1	SDSS DR5, FIRST, USNO	6 479	785	27 984
S2	SDSS DR5, ROSAT, USNO	6 360	2 357	5 333

Table 2 Mean Values of Parameters for S1

Parameters	Catalogs	Quasars	Stars	Galaxies
$u-g$	SDSS	0.56 ± 0.76	1.89 ± 0.79	1.85 ± 0.84
$g-r$	SDSS	0.30 ± 0.36	0.97 ± 0.46	1.17 ± 0.44
$r-i$	SDSS	0.19 ± 0.22	0.58 ± 0.45	0.53 ± 0.19
$i-z$	SDSS	0.11 ± 0.18	0.30 ± 0.26	0.34 ± 0.13
r	SDSS	18.70 ± 1.11	18.38 ± 1.45	17.32 ± 1.56
$\log(f_{\rm peak})$	FIRST	0.90 ± 0.72	0.49 ± 0.47	0.40 ± 0.38
$\log(f_{\rm int})$	FIRST	0.92 ± 0.75	0.58 ± 0.51	0.48 ± 0.43
$f_{\rm maj}$	FIRST	2.11 ± 2.19	3.80 ± 3.64	3.84 ± 3.36
$f_{\rm min}$	FIRST	0.64 ± 1.00	1.37 ± 2.13	1.10 ± 1.70
$f_{\rm pa}$	FIRST	98.01 ± 56.41	88.77 ± 54.47	89.44 ± 54.19
$z+2.5\log(f_{\rm peak})$	SDSS,FIRST	20.65 ± 2.16	18.71 ± 1.65	17.46 ± 1.67
$z+2.5\log(f_{\rm int})$	SDSS,FIRST	20.70 ± 2.23	18.94 ± 1.71	17.67 ± 1.72
$B-R$	USNO-B1.0	-0.48 ± 1.33	-0.64 ± 3.27	-0.82 ± 3.65

Table 3 Mean Values of Parameters for S2

Parameters	Catalogs	Quasars	Stars	Galaxies
$u-g$	SDSS	0.32 ± 0.41	1.68 ± 1.07	1.68 ± 0.90
$g-r$	SDSS	0.22 ± 0.29	0.90 ± 0.79	1.02 ± 0.47
$r-i$	SDSS	0.18 ± 0.21	0.51 ± 0.71	0.45 ± 0.30
$i-z$	SDSS	0.14 ± 0.21	0.25 ± 0.62	0.31 ± 0.29
r	SDSS	18.18 ± 1.08	18.58 ± 1.76	17.66 ± 1.65
log(CR)	RASS	-1.46 ± 0.33	-1.48 ± 0.34	-1.44 ± 0.39
HR1	RASS	-0.09 ± 0.53	0.05 ± 0.61	0.22 ± 0.58
HR2	RASS	0.11 ± 0.58	0.10 ± 0.56	0.17 ± 0.52
ext	RASS	3.77 ± 8.24	4.70 ± 11.44	8.96 ± 21.11
extl	RASS	0.46 ± 2.93	0.83 ± 5.83	9.14 ± 126.25
$u+2.5\log$(CR)	SDSS, RASS	15.06 ± 1.05	17.47 ± 2.70	16.76 ± 2.20
$B-R$	USNO-B1.0	-0.43 ± 1.13	-0.34 ± 3.24	-0.96 ± 2.67

$\log(f_{\rm peak})$, $\log(f_{\rm int})$, $f_{\rm maj}$, $f_{\rm min}$, $f_{\rm pa}$, while from the RASS survey are log(CR), HR1, HR2, ext, extl; those from the USNO-B1.0 catalog are the B and R magnitudes. For S1, the attributes that we used are $u-g$, $g-r$, $r-i$, $i-z$, r, $\log(f_{\rm peak})$, $\log(f_{\rm int})$, $f_{\rm maj}$, $f_{\rm min}$, $f_{\rm pa}$, $z+2.5\log(f_{\rm peak})$, $z+2.5\log(f_{\rm int})$, $B-R$. The attributes of $z+2.5\log(f_{\rm peak})$ and $z+2.5\log(f_{\rm int})$ can be viewed as radio-to-optical flux ratios. For S2, $u-g$, $g-r$, $r-i$, $i-z$, r, log(CR), HR1, HR2, ext, extl, $u+2.5\log$(CR), $B-R$ are considered. The attribute of $u+2.5\log$(CR) can be taken as an X-ray-to-optical flux ratio.

The statistical properties of S1 and S2 are summarized in Tables 2–3. From Tables 2–3, it is clear that the properties of galaxies are very similar to stars due to the fact that galaxies are made up of stars, but quasars show different traits from stars and galaxies. Tables 2–3 also indicate that the features selected as the input of classifiers are reasonable and applicable to discriminate quasars, stars and galaxies; especially for separating quasars from non-quasars.

4 RESULTS AND DISCUSSIONS

The random forest classifier is firstly applied to separately classify quasars from stars with S1 and S2. Since the random forest can compute the feature weight of each attribute, the method may be used for feature weight and feature selection. Therefore, we use it on S1 and S2, respectively. The weight values of each attribute for the two samples are shown in Tables 4 and 5. It is obvious that, for S1, the important attributes in sequence are $u-r$, $g-r$, $r-i$, $i-z$, r, $\log(f_{\rm peak})$, $\log(f_{\rm int})$, $\log(f_{\rm maj})$, $f_{\rm min}$, $f_{\rm pa}$, $z+2.5\log(f_{\rm peak})$, $z+2.5\log(f_{\rm int})$, $B-R$; while for S2, the attributes are $u-g$, $g-r$, $r-i$, $i-z$, r, log(CR), HR1, HR2, ext, extl, $u+2.5\log$(CR), $B-R$. In the following experiments, the feature weight is adopted from Tables 4–5, the feature selection is done by using a cutoff value which is 18.00 for S1 and 30.00 for S2, in other words, the features whose weight is larger than the cutoff value are kept. As

Table 4 Feature Weight for Each Feature of S1

Parameter	Feature weight
$u-g$	142.353
$g-r$	96.731
$r-i$	67.797
$i-z$	60.550
r	44.559
$\log(f_{\rm peak})$	36.147
$\log(f_{\rm int})$	30.680
$f_{\rm maj}$	20.998
$f_{\rm min}$	19.061
$f_{\rm pa}$	16.725
$z+2.5\log(f_{\rm peak})$	12.867
$z+2.5\log(f_{\rm int})$	10.668
$B-R$	1.535

Table 5 Feature Weight for Each Feature of S2

Parameter	Feature weight
$u-g$	196.366
$g-r$	81.819
$r-i$	59.084
$i-z$	48.853
r	43.924
log(CR)	35.813
HR1	29.164
HR2	18.859
ext	11.293
extl	8.649
$u+2.5\log$(CR)	3.871
$B-R$	1.241

Table 6 Classification Result of Quasars and Stars using S1

Sample	All	Feature	Weighted	Feature	Selected	Feature
classified ↓ known →	quasars	stars	quasars	stars	quasars	stars
quasars	6430	119	6431	118	6423	110
stars	49	666	48	667	56	675
Accuracy(%)	99.24±0.20	84.84±3.57	99.26±0.22	84.95±3.58	99.14±0.37	85.99±3.70
Total accuracy(%)	97.69±0.44		97.71±0.43		97.72±0.40	

a result, for the following feature selection situation, $u-r$, $g-r$, $r-i$, $i-z$, r, $\log(f_{\rm peak})$, $\log(f_{\rm int})$, $\log(f_{\rm maj})$ and $f_{\rm min}$ are selected for S1, while $u-g$, $g-r$, $r-i$, $i-z$, r and log(CR) are chosen for S2. Then, we classify quasars from stars with all features, the weighted features and the selected features by a 10-fold cross-validation, respectively. The experimental results are shown in Tables 6–7. For S1, the classification accuracy of quasars is above 99.00% for the three situations, while that of stars is better than 84.00%. The whole accuracy is superior to 97.60%. For S2, the classification accuracy of quasars is better than 98.00%, and that of stars is higher than 96.00%. The overall accuracy is more than 98.00%. Tables 6–7 indicate that the classification result with S2 is better than that with S1. That is to say, it is easier to discriminate quasars from stars based on optical and X-ray information than based on optical and radio information. Tables 6–7 also show that the performance with the weighted features or the selected features is slightly better than that with all features. This is due to the fact that the random forest model builder is able to target the most useful variables. As a result, if we have many input features, we generally do not need to do any feature selection before we begin model building.

Secondly, the random forest classifier is utilized to discriminate quasars, stars and galaxies in S1 and S2, respectively. Based on the experimental results of Tables 6–7, the overall accuracy is nearly the same for all three situations, so we only consider the situation with all the features used for classification. The classification result of quasars, stars and galaxies with all the features is given in Tables 8–9. For all kinds of objects, the accuracy of stars is the lowest, which may be due to stars composing of the smallest sample. The overall accuracy adds up to 96.17% and 91.08% for S1 and S2, respectively. Clearly, the performance in Table 8 is superior to that in Table 9, which means the information from optical and radio bands is more useful for separating quasars, stars and galaxies than that from optical and X-ray bands.

The random forest classifier (Breiman 2001) applies random feature selection in the tree induction process. Random forests generally exhibit a substantial performance improvement over single tree classifiers such as CART and C4.5. It yields generalization error rates that compare favorably to Adaboost, yet is more robust to noise. Only in terms of accuracy, Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) are slightly superior to Random Forests (Zhao & Zhang 2007). However, for the ANN approach, the optimal architecture is not easy to obtain; moreover, it is inclined to become stuck in local minima during the training stage. Although SVMs do not require the architecture to be

Table 7 Classification Result of Quasars and Stars using S2

Sample	All	Feature	Weighted	Feature	Selected	Feature
Classified ↓ known →	quasars	stars	quasars	stars	quasars	stars
quasars	6281	87	6281	87	6297	86
stars	79	2270	79	2270	63	2271
Accuracy(%)	98.76±0.61	96.31±0.99	98.76±0.61	96.31±0.99	99.01±0.60	96.35±1.07
Total accuracy(%)	98.10±0.51		98.10±0.51		98.29±0.52	

Table 8 Classification Result of Quasars, Stars and Galaxies using S1

Classified ↓ Known →	Quasars	Stars	Galaxies
quasars	5783	70	364
stars	15	516	33
galaxies	681	199	27587
Accuracy(%)	89.26±1.47	65.732±4.79	98.58±1.65
Total accuracy(%)	96.17±0.30		

Table 9 Classification Result of Quasars, Stars and Galaxies using S2

Classified ↓ Known→	Quasars	Stars	Galaxies
quasars	6005	67	409
stars	40	2055	188
galaxies	315	235	4736
Accuracy(%)	94.42±1.12	87.19±2.54	88.81±1.57
Total accuracy(%)	91.08±0.92		

chosen before training, the optimal parameters in the models can be obtained only with much effort. In addition, the ANN method shows its inferiority when faced with high dimensional data. For Random Forests, we need not consider the dimensionality of data; this algorithm can automatically select the important attributes used for classification or regression. It picks out the important attributes and keeps the original physical meaning of the attributes. Thus, astronomers conveniently know which attributes make more contributions to a data set. Since the random forest method can implement feature selection, it may be naturally used for data preprocessing when other classifiers (e.g. ANNs and SVMs) are employed.

5 CONCLUSIONS

The random forest algorithm was investigated for classification of multiwavelength data sets, which is both a challenging and important classification problem in astronomy. Moreover, one of the most important tasks in astronomy is to classify quasars, stars and galaxies; especially for discriminating quasars from stars. In experiments, the random forest classifier performed well in the classification of quasars, stars and galaxies as well as the separation of quasars from stars with data from optical and X-ray bands or with data from optical and radio bands. While separating quasars from stars, the results showed that the data from optical and X-ray bands outperformed that of the data from optical and radio bands. Otherwise, when classifying quasars, stars and galaxies, the performance based on the information of optical and radio bands showed its superiority. In terms of accuracies, the accuracy with the selected features or with the weighted features was better or comparable to the accuracies obtained by all features. Unlike artificial neural networks (ANNs), the random forest algorithm does not overfit, and it does not require guidance. Moreover, the algorithm can estimate the importance of features used for classification. Such estimation can be applied for feature extraction and/or feature weighting in multisource data classification where it can reduce the dimensionality. When the random forest method is applied for feature selection and other algorithms as the classifier, it can improve the efficiency of these algorithms. Since Random Forests have so many advantages, they will be an effective tool in astronomy which can be used for classification and feature selection/weighting. The random forest classifier created by known samples should predict the type of unknown objects and choose the sources of interest for followup observation. Random Forests can also detect outliers, which can be very useful when some of the cases may be rare or mislabeled. This is very important for astronomers who are eager to find rare or unknown objects. As a result, the random forest algorithm will be an interesting and applicable tool in astronomy.

Acknowledgements We are very grateful to the referee's constructive and insightful suggestions. This paper is funded by the National Natural Science Foundation of China under grant under Grant Nos. 10473013, 90412016 and 10778724, and by the 863 project under Grant No. 2006AA01A120. This research has made use of data products from the SDSS survey. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

References

Adams, A., & Woolley, A. 1994, Vistas in Astronomy, 38, 273

Albert, J., et al. 2008, Nucler Instruments and Methods in Physics Research A, 588, 424 (astro-ph/0709.3719)

Bailey, S., Aragon, C., Romano, R., et al. 2007, ApJ, 665, 1246

Becker, R. H., Helfand, D. J., White, R. L., et al. 1997, ApJ, 475, 479

Breiman, L. 2001, Machine Learning, 45, 5

Breiman, L., Last, M., & Rice, J. in Statistical challenges in astronomy, Third Statistical Challenges in Modern Astronomy (SCMA III) Conference, University Park, PA, USA, July 18-21 2001, eds., Eric D. Feigelson, G. Jogesh Babu, (New York: Springer), 2003, 243

Carliles, S., Budavári, T., Heinis, S., et al. 2007, astro-ph/0711.2477

Monet, D. G., Levine, S. E., Casian, B., et al. 2003, A&A, 125, 984

Schlegel, D. J., Finkbeiner, D. P., & Davix, M. 1998, ApJ, 500, 525

Sodré, L. Jr., & Cuevas, H. 1994, Vistas in Astronomy, 38, 286

Storrie-Lombardi, M. C., Lahav, O., Sodre, L. Jr., et al. 1992, MNRAS, 259, 8

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

York, D. G., Adelman, J., Anderson, J. E. Jr., et al. 2000, AJ, 120, 1579

Zhang, Y., & Zhao, Y. 2003, ChJAA (Chin. J. Astron. Astrophys.), 3, 183

Zhang, Y., & Zhao, Y. 2007, ChJAA (Chin. J. Astron. Astrophys.), 7, 289

Zhang, Y., & Zhao, Y. 2003, PASP, 115, 1006

Zhang, Y., & Zhao, Y. 2004, A&A, 422, 1113

Zhang, Y., & Zhao, Y. 2008, ADASS, in press

Zhao, Y., & Zhang, Y. 2007, Advances in Space Research, 41, 1955