Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 2289–2303 (2022) https://doi.org/10.1093/mnras/stab3165
Advance Access publication 2021 November 10

Photometric redshift estimation of BASS DR3 quasars by machine learning

Changhua Li,[1,2,3] Yanxia Zhang,[1,4★] Chenzhou Cui,[1,3★] Dongwei Fan,[1,3] Yongheng Zhao,[1] Xue-Bing Wu,[5,6] Jing-Yi Zhang,[1,4] Jun Han,[1,3] Yunfei Xu,[1,3] Yihan Tao,[1,3] Shanshan Li[1,2,3] and Boliang He[1,2,3]

[1]*National Astronomical Observatories, Beijing 100101, China*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*National Astronomical Data Center, Beijing 100101, China*
[4]*CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, China*
[5]*Department of Astronomy, School of Physics, Peking University, Beijing 100871, China*
[6]*Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China*

Accepted 2021 October 27. Received 2021 October 26; in original form 2021 September 22

ABSTRACT

Correlating Beijing–Arizona Sky Survey (BASS) data release 3 (DR3) catalogue with the ALLWISE data base, the data from optical and infrared information are obtained. The quasars from Sloan Digital Sky Survey are taken as training and test samples while those from LAMOST are considered as external test sample. We propose two schemes to construct the redshift estimation models with XGBoost, CatBoost, and Random Forest. One scheme (namely one-step model) is to predict photometric redshifts directly based on the optimal models created by these three algorithms; the other scheme (namely two-step model) is to first classify the data into low- and high-redshift data sets, and then predict photometric redshifts of these two data sets separately. For one-step model, the performance of these three algorithms on photometric redshift estimation is compared with different training samples, and CatBoost is superior to XGBoost and Random Forest. For two-step model, the performances of these three algorithms on the classification of low and high redshift subsamples are compared, and CatBoost still shows the best performance. Therefore, CatBoost is regarded as the core algorithm of classification and regression in two-step model. In contrast to one-step model, two-step model is optimal when predicting photometric redshift of quasars, especially for high-redshift quasars. Finally, the two models are applied to predict photometric redshifts of all quasar candidates of BASS DR3. The number of high-redshift quasar candidates is 3938 (redshift ≥3.5) and 121 (redshift ≥4.5) by two-step model. The predicted result will be helpful for quasar research and follow-up observation of high-redshift quasars.

Key words: methods: statistical – techniques: photometric – astronomical data bases: miscellaneous – galaxies: distances and redshifts – quasars: general – galaxies: photometric.

1 INTRODUCTION

Redshift is an important feature of celestial objects and reflects the distance between celestial objects and the Earth. By the redshift, the distance between celestial objects and the Earth can be measured, which is of great significance for the research about spatial position, formation and evolution, and luminosity function of celestial objects. In general, the redshifts of celestial objects can be estimated more accurately through the spectra of celestial objects. However, the spectral observation of large-scale celestial objects is a time-consuming task, and especially for faint sources, it is almost impossible to get their spectra at present. With the construction and operation of many large sky survey observation equipment, a great amount of multiband photometric data are obtained. Therefore, the study of redshift estimation of celestial objects through photometric data has great significance. Although we cannot directly get the redshifts of celestial objects from photometric data, the relationship between photometric data and redshifts of the celestial objects can be reflected through a specific algorithm. Baum (1957) and Koo (1985) proposed methods to estimate the redshifts of galaxies based on photometric data. Their results show that the redshifts of celestial objects can be measured well through multiband photometric data.

In recent years, machine learning methods have been widely used in photometric redshift estimation: for instance, *k*-Nearest Neighbours (*k*NN; Ball et al. 2007; Zhang et al. 2013), Gaussian process regression (Way & Srivastava 2006; Way et al. 2009; Bonfield et al. 2010), kernel regression (Wang et al. 2007), Self-Organizing Map (Way & Klose 2012; Carrasco & Brunner 2014), Support Vector Machine (Jones & Singal 2017; Schindler et al. 2017; Jin et al. 2019), Random Forest (RF; Carliles et al. 2010; Schindler et al. 2017), Artificial Neural Networks (Firth, Lahav & Somerville 2003; Zhang, Li & Zhao 2009; Yè et al. 2010; Cavuoti et al. 2012, 2017; Brescia et al. 2013), XGBoost (Jin et al. 2019), and deep learning (Curran, Moss & Perrott 2021). Moreover, researchers continuously try to develop new algorithms, improve old methods, or make innovations

★ E-mail: zyx@bao.ac.cn (YZ); ccz@bao.ac.cn (CC)

Table 1. Parameters in the known samples.

Parameters	Definition	Catalogue	Waveband
id	Source ID	BASS	–
RA	Right ascension in decimal degrees	BASS	–
Dec.	Declination in decimal degrees	BASS	–
gKronMag	Kron magnitude in the *g* band	BASS	Optical band
rKronMag	Kron magnitude in the *r* band	BASS	Optical band
zKronMag	Kron magnitude in the *z* band	BASS	Optical band
gPSFMag	PSF magnitude in the *g* band	BASS	Optical band
rPSFMag	PSF magnitude in the *r* band	BASS	Optical band
zPSFMag	PSF magnitude in the *z* band	BASS	Optical band
g	Extinction-corrected PSF magnitude in the *g* band	BASS	Optical band
r	Extinction-corrected PSF magnitude in the *r* band	BASS	Optical band
z	Extinction-corrected PSF magnitude in the *z* band	BASS	Optical band
W1mag	*W1* magnitude	ALLWISE	Infrared band
W2mag	*W2* magnitude	ALLWISE	Infrared band
W1	Extinction-corrected *W1* magnitude	ALLWISE	Infrared band
W2	Extinction-corrected *W2* magnitude	ALLWISE	Infrared band
Redshift	Spectral redshift	SDSS, LAMOST	–

in algorithm applications. Hoyle (2016) applied deep neural networks to estimate photometric redshifts of galaxies by using the full galaxy image in each measured band. Leistedt & Hogg (2017) presented a new method to infer photometric redshifts in deep galaxy and quasar surveys, which combined the advantages of both machine learning methods and template fitting methods by building template spectral energy distributions directly from the spectroscopic training data. Zhang et al. (2019) put forward a new strategy for photometric redshift estimation of quasars. Han et al. (2021) devised a new approach GeneticKNN based on *k*NN and genetic algorithm for photometric redshift estimation of quasars.

Although a large number of algorithms have been used in this field, there is still large room for improvement. Moreover, due to the continuous increase in the amount of photometric data, it is necessary to speed up training and predicting while improving accuracy. In this paper, we explore three methods (CatBoost, XGBoost, and RF) to estimate photometric redshifts of quasars and then compare two schemes (one- and two-step models) for photometric redshift estimation of quasars. The sample used for this issue is described in Section 2. Then, the adopted methods are briefly introduced in Section 3. Based on the samples, the different schemes for photometric redshift estimation of quasars by CatBoost, XGBoost, and RF are depicted in detail and compared in Section 4. The introduction and application of the two-step model are presented in Section 5. Finally, we summarize the results of this paper in Section 6.

2 DATA

The Beijing–Arizona Sky Survey (BASS; Zou et al. 2017a, b) and MOSAIC *z*-band Legacy Survey (MzLS; Silva et al. 2016) are optical imaging surveys to provide galaxy and quasar targets for follow-up observation by the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration 2016). They survey the northern Galactic cap at $\delta > 30^\circ$ and cover about 5400 deg^2. The BASS DR3 was released in 2019, which contains the data from all BASS and MzLS observations from 2015 January to 2019 March (Zou et al. 2019). The DR3 includes single-epoch photometric catalogue and co-added photometric catalogue. In this paper, we used co-added photometric catalogue from BASS DR3, which can be downloaded from https://nadc.china-vo.org/data/data/bassdr3coadd/f.

The Sloan Digital Sky Survey (SDSS; York et al. 2000) has been conducting sky survey for about 20 yr, acquiring a large amount of spectral and photometric data. The DR16 quasar catalogue (DR16Q) from SDSS includes 750 414 quasars (Blanton et al. 2017).

The Large Sky Area Multi-object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012; Luo et al. 2015) may observe 4000 spectra in an observation to a limiting magnitude as faint as $r = 19$ at the resolution $R = 1800$. The first-phase sky survey in 5 yr has been finished. The number of quasars in the fifth data release (DR5; http://dr5.lamost.org/) adds up to 52 453 quasars.

The *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) is an all-sky survey project in mid-infrared band. On the basis of the *WISE* work, AllWISE has obtained better data than *WISE* on photometric sensitivity and accuracy as well as astrometric precision.

For simplicity, we use the released catalogue in Li et al. (2021), which provides optical information from BASS DR3 and infrared information from ALLWISE. From this catalogue, we select all possible quasar candidates when one of *Class_b*, *Class_bi*, *Class_m*, and *Class_mi* is zero; the total number is 26 200 778. These selected sources are cross-matched with SDSS DR16Q and LAMOST DR5 quasars in 2 arcsec radius, respectively. Then, we obtain known samples BS_W and BL_W with spectral redshifts, respectively, from SDSS DR16Q and LAMOST DR5. For the known samples, we extinction correct all photometries and use AB magnitudes referring to the work (Schindler et al. 2017). The parameters about known samples are extracted from BASS, ALLWISE, SDSS, and LAMOST data bases, as described in Table 1. The spectroscopic redshift distributions of the known samples are shown in Fig. 1.

Then, we split the sample BS_W into two data sets BSO and BSW according to whether both *W1mag* and *W2mag* are *NULL* or not. When both *W1mag* and *W2mag* are *NULL*, this source belongs to the sample BSO; otherwise, it belongs to the sample BSW. Correspondingly, the sample BL_W is divided into two data sets BLO and BLW. In this paper, the samples BS_W and BSW are used as training set and test set, while BL_W and BLW are used as external test set.

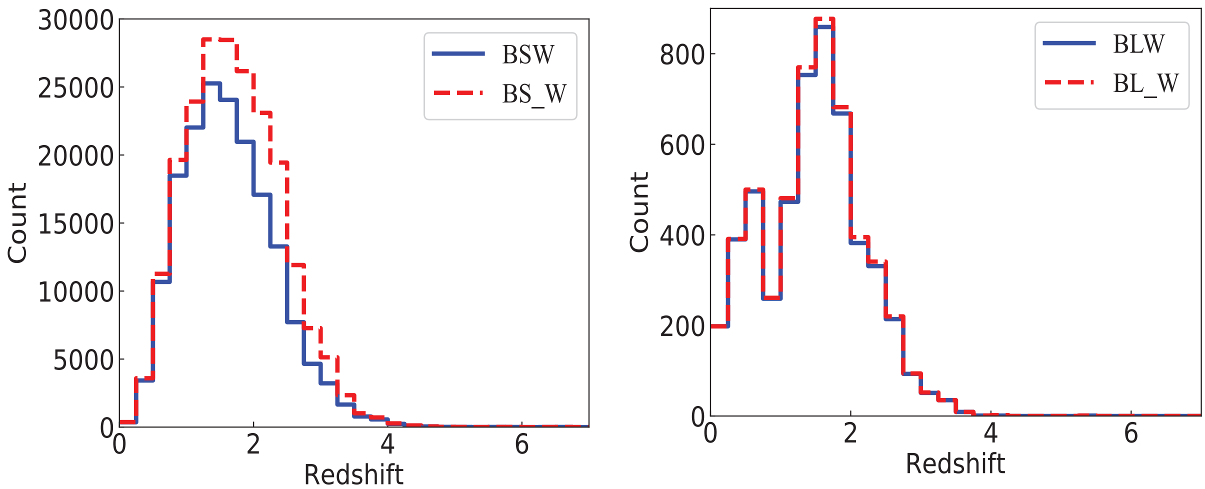


Figure 1. Left-hand panel: the distribution of spectroscopic redshifts for known samples BSW (blue line) and BS_W (red dash line); right-hand panel: the distribution of spectroscopic redshifts for known samples BLW (blue line) and BL_W (red dash line).

3 METHOD

3.1 CatBoost

Gradient Boosted Decision Trees (GBDT; Friedman 2001) are a powerful tool for classification and regression tasks. CatBoost is developed by Yandex researchers and engineers (Dorogush, Ershov & Yandex 2018), and it is a high-performance open-source algorithm. It is a member of the family of GBDT machine learning ensemble techniques, which applies boosting method to build strong classifiers by means of learning multiple weak classifiers or regressors. It uses Oblivious Decision Trees (ODT) to build decision trees, which are full binary trees; furthermore, all non-leaf nodes of ODT will have the same splitting criteria. This design is helpful to speed up the score and avoid overfitting. It supports categorical features and text features in data without additional pre-processing, and uses the ordered target statistics method for encoding new features. Besides, it is easy to get good result with the default set of model parameters. Thus, a lot of time for tuning its model parameters will be saved. It also has fast and scalable GPU version fit for handling big data. Compared to GBDT, it has two innovations: ordered target statistics and ordered boosting.

3.2 XGBoost

XGBoost (Chen & Guestrin 2016) is an excellent ensemble learning algorithm. Compared with other ensemble models, it can improve the model's robustness by introducing regular terms and column sampling; also, when each tree selects the split point, a parallelization strategy will be adopted to improve the model's running speed. Besides, XGBoost can overcome the limitations of computational speed and accuracy to a certain extent, requires less training and prediction time, and supports various objective functions, when performing classification and regression tasks.

3.3 Random Forest

RF (Breiman 2001) is based on bagging models built using the decision tree method. RF uses bootstrap resampling technology to randomly select K subsamples from the original training sample with replacement to generate a new training sample set, and build K classification trees to form an RF. Each tree in the forest has the same distribution, and the classification error depends on the classification ability of each tree and the correlation between them. Feature selection uses a random method to split each node, and then compares the errors generated in different situations. Therefore, RF uses the average to improve the predictive accuracy and control overfitting. In the scikit-learn, the implementation of RF combines classifiers by averaging their probabilistic prediction, instead of letting each classifier vote for a single class.

XGBoost and RF methods have wide applications in astronomy, such as classification of unknown source (Mirabal et al. 2016), quasar candidate selection (Jin et al. 2019), photometric redshift estimation (Zhang et al. 2019), etc. CatBoost has various applications in different fields, such as finance, investment, petroleum, and medicine. In this paper, we use XGBoost, CatBoost, and RF as supervised learning algorithms to build regressors for photometric redshift estimation of different quasar samples with different features. We also compare the performances of these three algorithms when classifying quasar sample into low and high redshift subsamples. XGboost and RF PYTHON packages are provided by scikit-learn (Pedregosa et al. 2011) and all computing runs in the cloud computing environment of National Astronomical Data Centre (NADC; Li et al. 2017).

4 PHOTOMETRIC REDSHIFT ESTIMATION

4.1 Regression metrics

The performance evaluation of different algorithms about photometric redshift estimation depends on different metrics, such as the residual between the spectroscopic and photometric redshifts, $\Delta z = z_{\rm spec} - z_{\rm photo}$, the mean absolute error (MAE), and the mean squared error (MSE). They are defined as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=0}^{n-1} |z_i - \widehat{z}_i| \tag{1}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n-1} (z_i - \widehat{z}_i)^2, \tag{2}$$

where z_i is the true redshift, $\widehat{z}_i$ is the predicted redshift value, and n is the sample size.

The fraction of test sample that satisfies $|\Delta z| < e$ is usually used to evaluate the redshift estimation, where e is a given residual threshold (see Schindler et al. 2017, and references therein). In reality, the redshift normalized residual [Δz(norm)] is often adopted, and we use $e = 0.3$.

$$\Delta z(\mathrm{norm}) = \frac{z_{\mathrm{spec}} - z_{\mathrm{phot}}}{1 + z_{\mathrm{spec}}} \tag{3}$$

$$\delta_{0.3} = \frac{N_{|\Delta z(\mathrm{norm})|<0.3}}{N_{\mathrm{total}}} \tag{4}$$

Then, we use five additional metrics defined below as a reference for the performance evaluation of different machine learning methods: R^2, bias (the average separation between prediction and true values), the standard deviation between the photometric redshifts and the spectroscopic redshifts, the normalized median absolute deviation (σ_{NMAD}), and the outlier fraction (O) (Ben et al. 2021; Curran et al. 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} |(z_i - \widehat{z}_i)^2}{\sum_{i=1}^{n} (z_i - \overline{z})^2} \tag{5}$$

$$\mathrm{Bias} = < z_{\mathrm{spec}} - z_{\mathrm{phot}} > \tag{6}$$

$$\sigma_{\Delta z} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1} (\Delta z)^2} \tag{7}$$

$$\sigma_{\mathrm{NMAD}} = 1.48 \times \mathrm{median}\left|\frac{z_{\mathrm{spec}} - z_{\mathrm{phot}}}{1 + z_{\mathrm{spec}}}\right| \tag{8}$$

$$\text{Outlier fraction } (O) = \frac{N_{|\Delta z(\mathrm{norm})|>0.15}}{N_{\mathrm{total}}} \tag{9}$$

4.2 Feature selection

When handling high-dimensional data, feature selection is the key factor influencing the performance of a machine learning algorithm. Feature selection not only reduces the dimension of data and rules out unimportant features, but also contributes to improve the accuracy of an algorithm. During data pre-processing, we need to obtain optimal features first, and adopt two steps for feature selection.

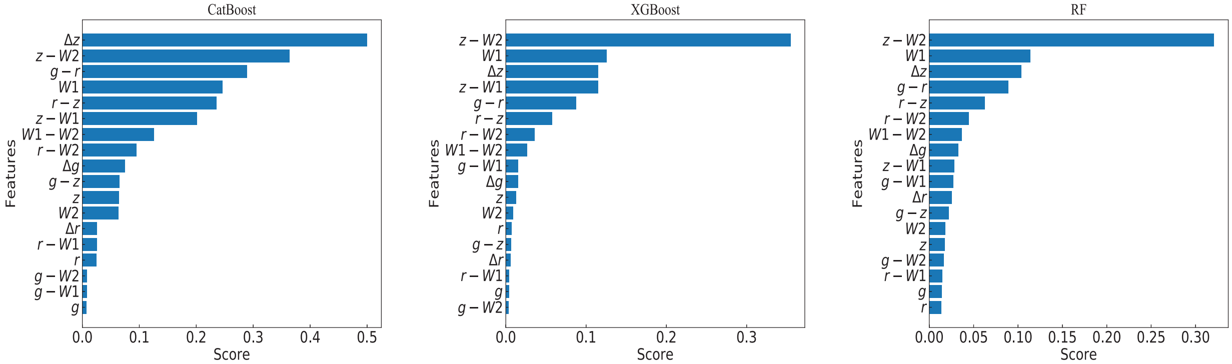
According to the parameters listed in Table 1, we define optical and infrared features. All optical features include *KronMag* − *PSFMag* of g, r, z (namely Δg, Δr, Δz, respectively), g, r, z, $g - r$, $r - z$, and $g - z$, and all infrared features are $W1$, $W2$, $g - W1$, $r - W1$, $z - W1$, $g - W2$, $r - W2$, $z - W2$, and $W1 - W2$. For the samples BSW and BS_W, they have all optical and infrared features. The difference between the samples BSW and BS_W is that the sample BS_W includes the samples BSW and BSO. Each source in the sample BSW contains both optical and infrared information, while the sample BSO only includes optical information. The difference between the samples BLW and BL_W is same. The quasars of BSW and BS_W are from SDSS DR16Q, and the quasars of BLW and BL_W are from LAMOST DR5. As for the samples BSO and BLO, only the features from optical band are used. The sources in the samples BS_W and BL_W have no match sources in ALLWISE and no infrared features, so the features of these sources related to infrared band are missing. In other words, the samples BS_W and BL_W contain missing values. We set all missing values to 0 when using RF. XGBoost and CatBoost support missing values.

First, all features of training samples are evaluated by RF, XGBoost, and CatBoost methods, which can give the importance score of each feature. The importance type (*importance_type*) is *total_gain*. We sort these features by the importance score. The feature importance for different samples by different methods is shown in Fig. 2. Fig. 2 indicates that the feature importance is closely related to samples and algorithms. Furthermore, from the feature importance rank, it is seen that the infrared information is very important to the redshift estimation of quasars.

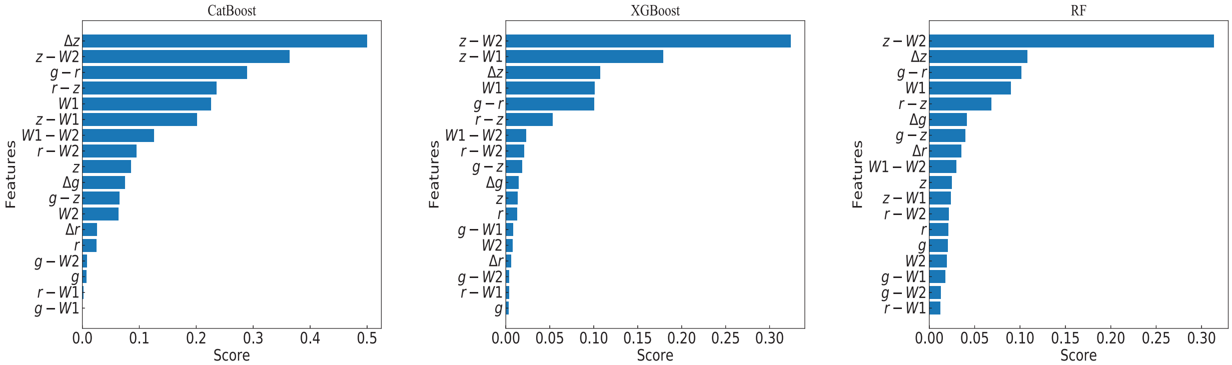
Secondly, when a method is adopted, according to the above rank of features, we select the top four features as initial input pattern to train a model, then add one feature in turn to the input pattern for training, and record all performance. MSE with different input features for different methods is described in Fig. 3. For simplicity, the best input pattern is adopted when MSE achieves the minimum. As shown in the left-hand panel of Fig. 3, for the sample BSW, the best input pattern is $z - W2$, $W1$, Δz, $z - W1$, $g - r$, $r - z$, $r - W2$, $W1 - W2$, $g - W1$, Δg, and z (11 features) by XGBoost; the best input pattern is Δz, $z - W2$, $g - r$, $W1$, $r - z$, $z - W1$, $W1 - W2$, $r - W2$, Δg, $g - z$, z, $W2$, and Δr (13 features) by CatBoost; and the best input pattern is $z - W2$, $W1$, Δz, $g - r$, $r - z$, $r - W2$, $W1 - W2$, Δg, $z - W1$, $g - W1$, Δr, $g - z$, $W2$, z, and $g - W2$ (15 features) by RF. As plotted in the middle panel of Fig. 3, for the sample BS_W, the optimal input pattern of XGBoost is $z - W2$, $z - W1$, Δz, $W1$, $g - r$, $r - z$, $W1 - W2$, $r - W2$, $g - z$, Δg, z, r, $g - W1$, $W2$, Δr, and $g - W2$ (16 features); the optimal input pattern of CatBoost is Δz, $z - W2$, $g - r$, $r - z$, $W1$, $z - W1$, $W1 - W2$, $r - W2$, z, Δg, $g - z$, $W2$, Δr, r, and $g - W2$ (15 features); and the optimal input pattern of RF is $z - W2$, Δz, $g - r$, $W1$, $r - z$, Δg, $g - z$, Δr, $W1 - W2$, z, $z - W1$, and $r - W2$ (12 features). As indicated in the right-hand panel of Fig. 3, for the sample BS_W only with optical features, the optimal input patterns of the three methods keep all optical features (nine features) and retain their order. With the best input patterns for different samples by the three methods, we adopt the default hyperparameters of models and fivefold cross-validation to get the average values of MSE, MAE, bias, σ_{NMAD}, $\sigma_{\Delta z}$, R^2, $\delta_{0.3}$, outlier fraction (O), and running *time*, which are shown in Table 2. Comparing all metrics in Table 2, CatBoost achieves better performance than XGBoost and RF for most metrics. In general, the better performance with optical and infrared features is achieved than only with optical features. Table 2 further confirms this fact.

4.3 Model parameter optimization

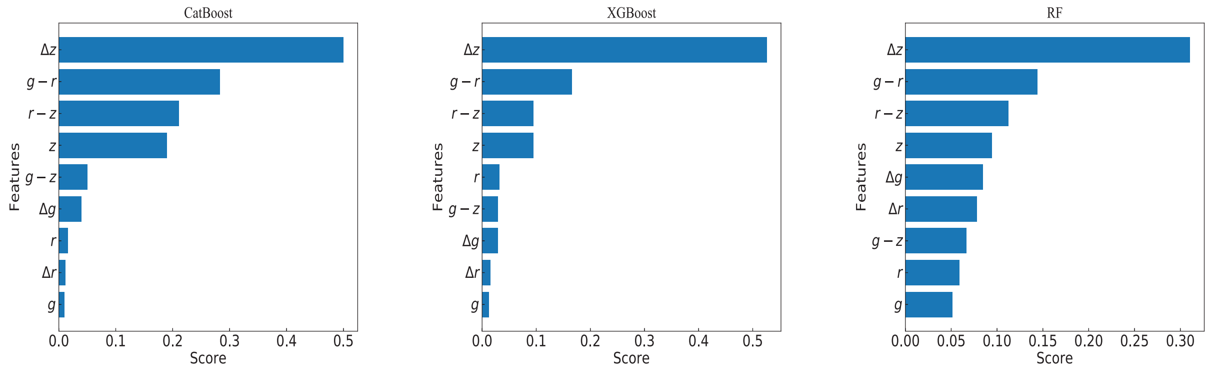
Since the optimal input patterns have been set, the next task is to determine the hyperparameters of models. Model parameter optimization is a very complex task. In order to reduce computing scale, we only choose some main hyperparameters for each method. For XGBoost, the important model parameters contain the maximum depth of individual trees (max_depth) and the number of weak estimators (n_estimators); for CatBoost, the key model parameters are the maximum depth of individual tree (depth) and the maximum number of trees (iterations); for RF, the main model parameters are the maximum depth of individual trees (max_depth) and the number of trees in the forest (n_estimators). We adopt the grid search method to get the optimal model parameters and fivefold cross-validation to get the average values of MSE, MAE, bias, σ_{NMAD}, $\sigma_{\Delta z}$, R^2, $\delta_{0.3}$, O, and running *time*. The optimal model parameters and performance are listed in Table 3. As described in Table 3, the best performance is obtained on the sample BSW for these three methods; MSE, MAE, bias, σ_{NMAD}, $\sigma_{\Delta z}$, R^2, $\delta_{0.3}$, and O, respectively, amount to 0.1576, 0.2649, −0.0006, 0.0999, 0.3971, 0.6680, 93.78, and 20.45 for CatBoost; 0.1617, 0.2669, −0.0009, 0.1001, 0.4022, 0.6595, 93.61, and 20.58 for XGBoost; 0.1626, 0.2686, 0.001, 0.1007, 0.4032, 0.6578, 93.50, and 20.83 for RF, respectively. Table 3 also shows that for the sample BS_W, the performance with optical and infrared information is better than that with only optical



(a) For the sample BSW with optical and infrared features



(b) For the sample BS_W with optical and infrared features



(c) For the sample BS_W only with optical features

Figure 2. The feature importance for different samples by CatBoost, XGBoost, and RF.

information for any method. From comparison of the performances of XGBoost, CatBoost, and RF in Table 3, CatBoost shows its superiority.

Then, with optimal model parameters and 80:20 training–test split, we train again for the samples BSW and BS_W. The scatter figure and Δz(norm) distribution of estimated photometric redshifts and spectroscopic redshifts are shown in Fig. 4. Fig. 4 further proves that CatBoost outperforms XGBoost and RF.

Finally, we use the three optimal models to train regressors with the total samples, and apply the total samples and external samples as test samples separately to validate the regressors; the validation performance is shown in Table 4. Table 4 tells that when the regressors created with optical and infrared information are applied, the performance of CatBoost with Pattern II is MSE = 0.1059, MAE = 0.2223, bias = -1.6×10^{-5}, $\sigma_{\rm NMAD}$ = 0.0872, $\sigma_{\Delta z}$ = 0.3254, $R^2 = 0.7780$, $\delta_{0.3}$ = 96.01 per cent, and O = 15.79 per cent for the test sample BSW; the performance of CatBoost with Pattern II is MSE = 0.1239, MAE = 0.2134, bias = 0.0265, $\sigma_{\rm NMAD}$ = 0.0797, $\sigma_{\Delta z}$ = 0.3520, R^2 = 0.7585, $\delta_{0.3}$ = 94.50 per cent, and O = 15.11 per cent for the test sample BLW. For the test sample BL_W or BSO, the regressor with Pattern V is better than that with Pattern VII, which suggests that the regressor with Pattern V had better be applied when the sources to be predicted only contain optical information. For the test sample BLW or BSW, the regressor with Pattern II is superior to that with Pattern V. Therefore, the sources with optical and infrared information had better be predicted by the regressor with Pattern II. Table 4 further shows that the CatBoost regressors are effective to predict photometric redshifts of quasars.

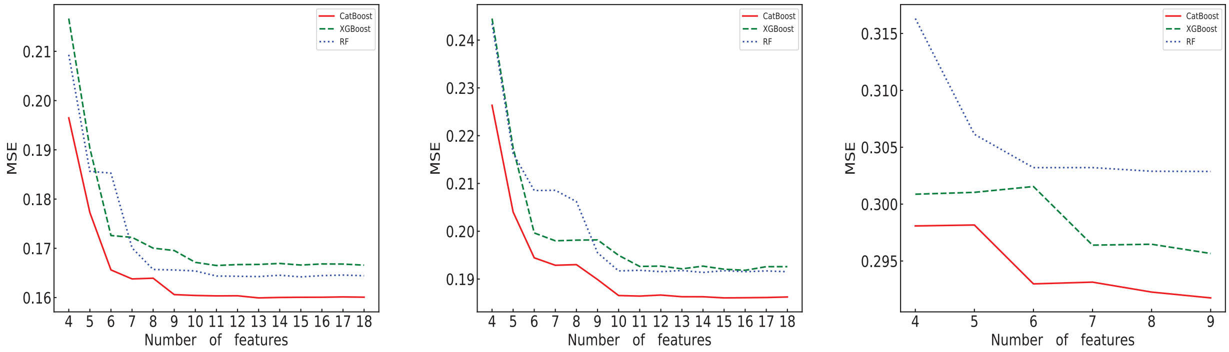


Figure 3. The performance of different methods with different input patterns for different samples. Left-hand panel: for the sample BSW with optical and infrared information; middle panel: for the sample BS_W with optical and infrared information; right-hand panel: for the sample BS_W only with optical information.

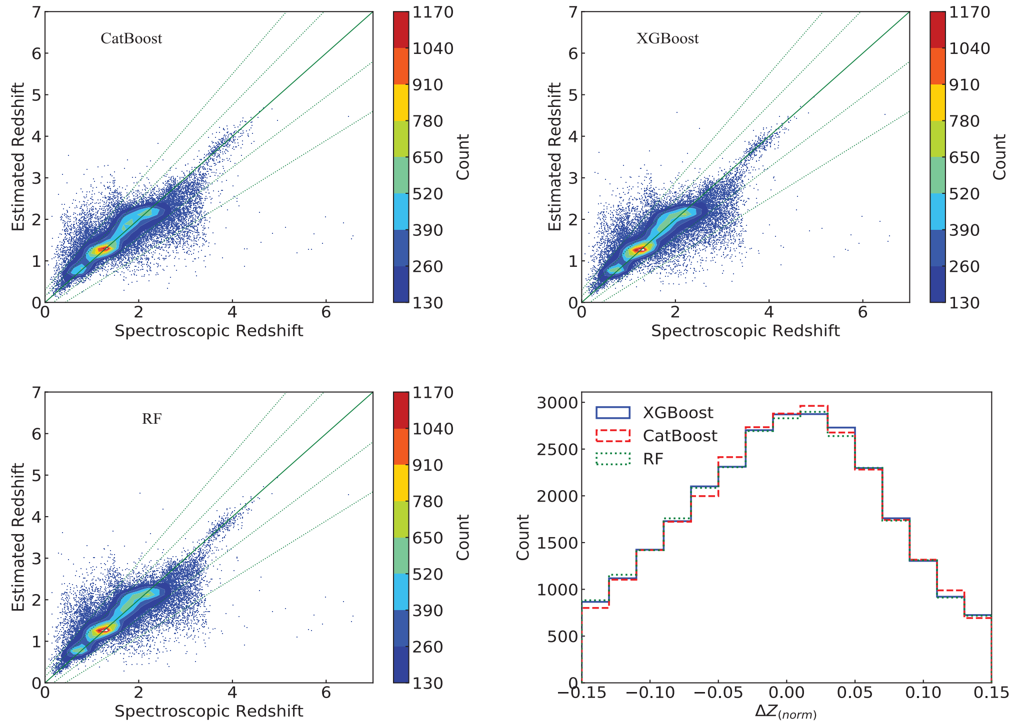
Table 2. The performance of photometric redshift estimation with optimal input features for each method with default model parameters.

Sample	Input pattern	Method	MSE	MAE	Bias	$\sigma_{\rm NMAD}$	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)	Time (s)
BSW	Pattern I	XGBoost	0.1666	0.2765	4.0×10^{-5}	0.1070	0.4082	0.6493	93.39	22.04	3
BSW	Pattern II	CatBoost	0.1600	0.2702	6.0×10^{-5}	0.1036	0.4000	0.6632	93.61	21.32	10
BSW	Pattern III	RF	0.1645	0.2700	2.2×10^{-3}	0.1008	0.4056	0.6537	93.35	21.01	442
BS_W	Pattern IV	XGBoost	0.1925	0.3064	-7.6×10^{-6}	0.1171	0.4387	0.6071	92.37	25.22	6
BS_W	Pattern V	CatBoost	0.1867	0.3012	-4.8×10^{-6}	0.1146	0.4321	0.6189	92.56	24.48	12
BS_W	Pattern VI	RF	0.1924	0.3009	-2.0×10^{-5}	0.1107	0.4387	0.6074	92.32	24.03	447
BS_W	Pattern VII	XGBoost	0.2959	0.4009	-1.0×10^{-4}	0.1657	0.5439	0.4008	87.27	37.40	4
BS_W	Pattern VII	CatBoost	0.2917	0.3988	-9.0×10^{-5}	0.1651	0.5403	0.4087	87.39	37.74	12
BS_W	Pattern VII	RF	0.3032	0.4045	8.7×10^{-3}	0.1648	0.5507	0.3858	86.52	37.93	200

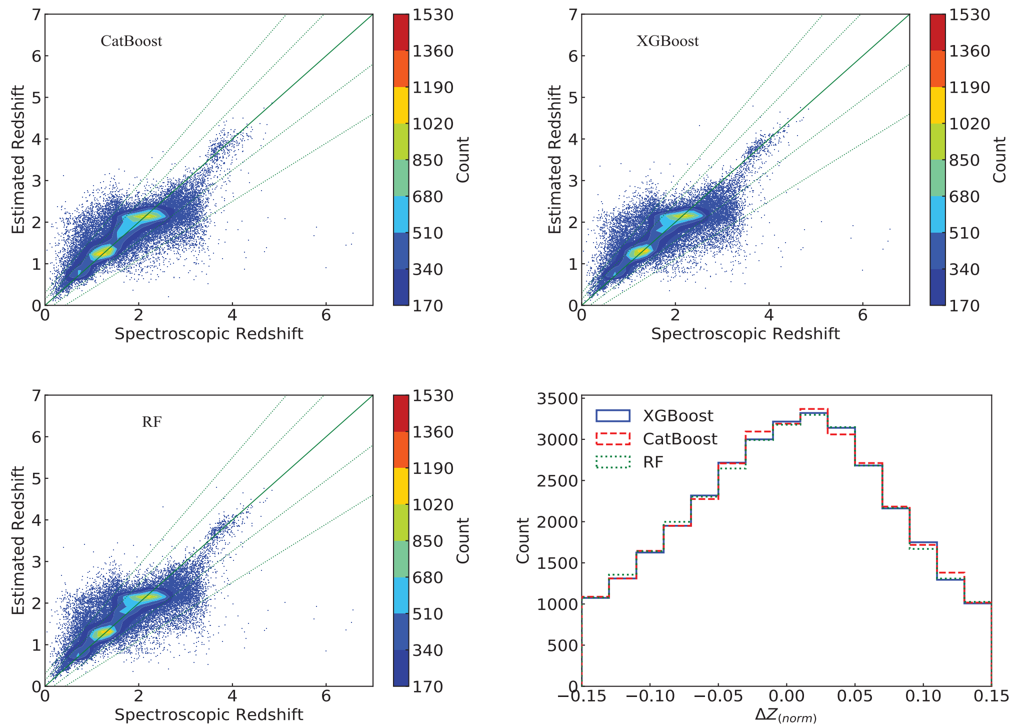
Notes. [a]Pattern I represents $z - W2$, $W1$, Δz, $z - W1$, $g - r$, $r - z$, $r - W2$, $W1 - W2$, $g - W1$, Δg, and z (11 features).
[b]Pattern II represents Δz, $z - W2$, $g - r$, $W1$, $r - z$, $z - W1$, $W1 - W2$, $r - W2$, Δg, $g - z$, z, $W2$, and Δr (13 features).
[c]Pattern III represents $z - W2$, $W1$, Δz, $g - r$, $r - z$, $r - W2$, $W1 - W2$, Δg, $z - W1$, $g - W1$, Δr, $g - z$, $W2$, z, and $g - W2$ (15 features).
[d]Pattern IV represents $z - W2$, $z - W1$, Δz, $W1$, $g - r$, $r - z$, $W1 - W2$, $r - W2$, $g - z$, Δg, z, r, $g - W1$, $W2$, Δr, and $g - W2$ (16 features).
[e]Pattern V represents Δz, $z - W2$, $g - r$, $r - z$, $W1$, $z - W1$, $W1 - W2$, $r - W2$, z, Δg, $g - z$, $W2$, Δr, r, and $g - W2$ (15 features).
[f]Pattern VI represents $z - W2$, Δz, $g - r$, $W1$, $r - z$, Δg, $g - z$, Δr, $W1 - W2$, z, $z - W1$, and $r - W2$ (12 features).
[g]Pattern VII represents Δz, Δg, Δr, $g - r$, $g - z$, $r - z$, g, r, and z (9 features).
[h]Patterns I, II, III, IV, V, VI, and VII represent the same defination in the following of this paper.

Table 3. The performance of photometric redshift estimation with the best features and optimal model parameters.

Sample	Input pattern	Method	Model parameter	MSE	MAE	Bias	$\sigma_{\rm NMAD}$	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)	Time (s)
BSW	Pattern I	XGBoost	max_depth = 11 n_estimators = 1000	0.1617	0.2669	−0.0009	0.1001	0.4022	0.6595	93.61	20.58	110
BSW	Pattern II	CatBoost	depth = 12 iterations = 4000	0.1576	0.2649	−0.0006	0.0999	0.3971	0.6680	93.78	20.45	421
BSW	Pattern III	RF	max_depth = 15 n_estimators = 500	0.1626	0.2686	0.0001	0.1007	0.4032	0.6578	93.50	20.83	1336
BS_W	Pattern IV	XGBoost	max_depth = 10 n_estimators = 1200	0.1865	0.2969	0.0001	0.1106	0.4318	0.6195	92.62	23.72	325
BS_W	Pattern V	CatBoost	depth = 12 iterations = 4000	0.1848	0.2960	0.0001	0.1105	0.4299	0.6228	92.69	23.55	468
BS_W	Pattern VI	RF	max_depth = 15 n_estimators = 500	0.1893	0.2990	0.0009	0.1111	0.4351	0.6136	92.44	23.82	1431
BS_W	Pattern VII	XGBoost	max_depth = 10 n_estimators = 1000	0.2930	0.3967	−0.0001	0.1626	0.5413	0.4066	87.43	37.17	110
BS_W	Pattern VII	CatBoost	depth = 12 iterations = 2000	0.2906	0.3970	−0.0001	0.1637	0.5393	0.4110	87.47	37.38	254
BS_W	Pattern VII	RF	max_depth = 15 n_estimators = 500	0.2944	0.3975	0.0003	0.1623	0.5425	0.4038	87.36	37.07	1124



(a) For the sample BSW



(b) For the sample BS_W

Figure 4. The first three plots in (a) and (b) show the performance of photometric redshift estimation with CatBoost, XGBoost, and RF, respectively. The green line represents $\Delta z = 0$; the green dotted lines represent Δz(norm) $= \pm 0.15$, ± 0.3, separately. The last plots in (a) and (b) show the distribution of Δz(norm).

5 TWO-STEP MODEL

In the whole quasar samples, the number of high-redshift quasars (redshift $\geq$3.5) is relatively small compared to that of low-redshift quasars (redshift <3.5). In the sample BS_W, there are only 2238 high-redshift quasars among the 213 359 quasars, while in the sample BSW, there are only 1798 high-redshift quasars among 174 645 quasars. The sample BS_W is divided into high redshift subsample BS_W_H and low redshift subsample BS_W_L; similarly,

Table 4. The performance of CatBoost for photometric redshift estimation.

Training sample	Input pattern	Test sample	MSE	MAE	Bias	$\sigma_{\rm NMAD}$	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)
BSW	Pattern II	BSW	0.1059	0.2223	-1.6×10^{-5}	0.0872	0.3254	0.7780	96.01	15.79
BSW	Pattern II	BLW	0.1239	0.2134	0.0265	0.0797	0.3520	0.7585	94.50	15.11
BS_W	Pattern V	BSW	0.1115	0.2279	-4.0×10^{-5}	0.0889	0.3340	0.7661	95.77	16.42
BS_W	Pattern V	BS_W	0.1365	0.2578	-7.0×10^{-6}	0.0981	0.3695	0.7239	94.84	19.42
BS_W	Pattern V	BLW	0.1265	0.2167	0.0268	0.0808	0.3557	0.7534	94.48	15.61
BS_W	Pattern V	BL_W	0.1277	0.2188	0.0296	0.0817	0.3574	0.7502	94.41	15.87
BS_W	Pattern V	BSO	0.2490	0.3929	1.2×10^{-4}	0.1550	0.4991	0.3884	90.64	32.93
BS_W	Pattern VII	BSO	0.3267	0.4363	−0.1853	0.1672	0.5719	0.1970	90.98	36.97
BS_W	Pattern VII	BL_W	0.2596	0.3681	0.1223	0.1639	0.5095	0.4925	85.82	37.61

the sample BSW is split into BSW_H and BSW_L. Table 5 shows the performance of photometric redshift estimation for high redshift and low redshift subsamples based on the best CatBoost regressors in Table 4. As described in Table 5, the performance on the high redshift subsamples is much worse than that on the low redshift subsamples for the sample BS_W or BSW. Meanwhile, Fig. 4 indicates that many high-redshift quasars have been predicted as low-redshift quasars. For the sample BS_W, there are only 1765 quasars that are correctly predicted as high-redshift quasars among 2238 high-redshift quasars, taking up about 78 per cent, while for the sample BSW, 1492 quasars are correctly predicted as high-redshift quasars among 1798, occupying about 83 per cent. It is obvious that the regressor fit for low-redshift quasars is not fit for high-redshift quasars. Thus, we propose a two-step model to improve the performance of redshift estimation for quasars, especially for high-redshift quasars.

5.1 The first step of the two-step model

In order to improve the performance of photometric redshift estimation of quasars, we put forward a new scheme of first classification and second regression (i.e. two-step model) for photometric redshift estimation. The first step of two-step model is to construct a classifier to discriminate the whole quasars into low and high redshift subsamples. Based on the samples BSW and BS_W, we add a column 'label', which is 1 if redshift ≥3.5, otherwise, set to 0. We adopt four standard metrics (Accuracy, Precision, Recall, and F1_score) to evaluate the performance of one classifier. Accuracy (Accu.) is the ratio of the number of the correctly classified samples to the total number of the samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (10)$$

Here, TP and TN are correctly classified by the classifier. FP shows the negative sample classified as positive. FN represents the positive sample classified as negative.

Precision (Prec.) is the ratio of the true positive (negative) sample in all the samples that are classified as positive (negative). Recall (Rec.) is the ratio of correctly classified positive (negative) samples in all the true positive (negative) samples.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (11)$$

F1_score (F1) is a weighted average of Precision and Recall.

$$\text{F1\_score} = \frac{2 \times (\text{Prec.} \times \text{Rec.})}{\text{Prec.} + \text{Rec.}} \quad (12)$$

Then, we use CatBoost, XGBoost, and RF to construct binary classifiers; the optimal performance and hyperparameters of the three classifiers by fivefold validation are shown in Table 6. Table 6 shows that CatBoost is superior to XGBoost and RF when separating the sample BSW or BS_W into low and high redshift subsamples considering Accuracy and F1_score. Moreover, Recall (completeness) of CatBoost achieves the best for high-redshift quasars. In order to find more high-redshift quasars, we should keep high completeness at high redshift in the first step of two-step model.

In order to further test the performance of CatBoost, XGBoost, and RF classifiers, we use the whole quasar samples from SDSS as training and test samples. The self-validation classification performance of the three classifiers is shown in Table 7. It can be seen from Table 7 that the CatBoost classifier based on the sample BS_W has better performance than that based on the sample BSW when the sample BSW is taken as test sample. The better performance is that Accuracy reaches 99.99 per cent, F1 is 99.96 per cent for the high redshift subsample, and F1 is 99.99 per cent for the low redshift subsample. Therefore, the CatBoost classifier based on the sample BS_W is more suitable for the sample BSW. In other words, the CatBoost classifier created on the training sample BS_W had better be applied to classify the sources with optical information or those with optical and infrared information. When the BSW sample is taken as training and test samples, XGBoost achieves the best performance while CatBoost and RF obtain the same performance (all metrics above 96.27 per cent). In other two situations, CatBoost is superior to XGBoost and RF; moreover, the running time of CatBoost is the fastest. Therefore, CatBoost is adopted as the core algorithm of classification in the first step of the whole scheme.

5.2 The second step of the two-step model

High and low redshift subsamples of BSW and BS_W are trained to construct new regressors by CatBoost, XGBoost, and RF, respectively. The optimal model parameters and performance of all subsamples for these three methods are shown in Table 8. As shown in Table 8, the performance of both low and high redshift subsamples has been significantly improved, especially for high-redshift quasars. Taking the evaluation metrics of regression into account except bias, $\sigma_{\rm NMAD}$, and O in Table 8, CatBoost outperforms XGBoost and RF for any sample. Only given $\sigma_{\rm NMAD}$, CatBoost obtains the best performance for the samples BSW_L and BS_W_L. Only considering O, CatBoost shows its superiority for the samples BSW_H, BSW_L, and BS_W_L. As a result, CatBoost is taken as the core regression algorithm for the two-step model.

Table 5. The performance of photometric redshift estimation for high and low subsamples with the best CatBoost regressors in Table 4.

Training sample	Pattern	Test sample	MSE	MAE	Bias	σ_{NMAD}	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)
BS_W	Pattern V	BS_W_H	0.4079	0.2799	−0.1943	0.0359	0.6387	−1.4640	96.15	8.04
BS_W	Pattern V	BS_W_L	0.1336	0.2575	0.0021	0.0989	0.3655	0.7019	94.83	19.34
BSW	Pattern II	BSW_H	0.4074	0.2575	−0.1741	0.0318	0.6383	−1.1900	96.27	6.7
BSW	Pattern II	BSW_L	0.1027	0.2219	0.0018	0.0879	0.3205	0.7591	96.00	15.89

Table 6. The performance of different classifiers for high and low redshift subsamples.

					High redshift			Low redshift			
Sample	Input pattern	Method	Parameter	Accu. (per cent)	Prec. (per cent)	Rec. (per cent)	F1 (per cent)	Prec. (per cent)	Rec. (per cent)	F1 (per cent)	Time (s)
BSW	Pattern I	XGBoost	max_depth = 6 n_estimators = 200	99.78	92.48	86.10	89.18	99.86	99.92	99.89	124
BSW	Pattern II	CatBoost	depth = 6 iterations = 1000	**99.79**	92.33	**87.04**	**89.60**	**99.87**	99.93	**99.90**	19
BSW	Pattern III	RF	max_depth = 14 n_estimators = 500	99.78	**93.91**	84.54	88.98	99.84	**99.94**	99.89	190
BS_W	Pattern IV	XGBoost	max_depth = 6 n_estimators = 100	99.77	92.30	85.12	88.56	**99.84**	99.92	99.87	46
BS_W	Pattern V	CatBoost	depth = 12 iterations = 1000	**99.78**	92.87	**85.21**	**88.88**	**99.84**	99.93	**99.88**	130
BS_W	Pattern VI	RF	max_depth = 13 n_estimators = 100	99.77	**93.88**	82.89	88.04	99.82	**99.94**	99.87	72

Table 7. The performance of different classifiers with different training and test samples.

					High redshift			Low redshift			
Training sample	Method	Input pattern	Test sample	Accu. (per cent)	Prec. (per cent)	Rec. (per cent)	F1 (per cent)	Prec. (per cent)	Rec. (per cent)	F1 (per cent)	Time (s)
BSW	CatBoost	Pattern II	BSW	99.96	100	96.27	98.10	99.96	100	99.98	0.25
BS_W	CatBoost	Pattern V	BSW	99.99	100	99.94	99.97	99.99	100	100	0.27
BS_W	CatBoost	Pattern V	BS_W	99.99	100	99.91	99.95	99.99	100	100	0.36
BSW	XGBoost	Pattern I	BSW	100	100	100	100	100	100	100	2
BS_W	XGBoost	Pattern IV	BSW	99.97	99.60	98.00	98.79	99.98	99.99	99.99	1.5
BS_W	XGBoost	Pattern IV	BS_W	99.97	99.68	97.90	98.78	99.98	99.99	99.99	2
BSW	RF	Pattern III	BSW	99.96	100	96.27	98.10	99.96	100	99.98	11
BS_W	RF	Pattern VI	BSW	99.93	99.53	94.16	96.77	99.94	99.99	99.96	3
BS_W	RF	Pattern VI	BS_W	99.92	99.33	93.07	96.10	99.93	99.99	99.96	3

5.3 Comparison of two-step model with one-step model

For the two-step model, its first step is to classify the sample into low and high redshift subsamples; its second step is to create CatBoost regressors for low and high redshift subsamples. According to the above experiments, CatBoost is regarded as the core algorithms for both classification and regression. The CatBoost classifier is trained on the BS_W sample no matter whether the sources own infrared information, while the CatBoost regressors for low and high redshift subsamples are trained on the samples BS_W_L and BS_W_H for sources only with optical information and trained on the samples BSW_L and BSW_H with optical and infrared information. We use the samples BSW, BS_W, BLW, and BL_W to test models. Table 4 gives the performance of photometric redshift estimation by one-step model. In order to compare the performances of photometric redshift estimation between one-step model and two-step model, the test results of the two models with different test samples are indicated in Table 9. Taking the samples BSW and BS_W for example, the comparison of predicted photometric redshifts with spectroscopic redshifts is described in Fig. 5 and the Δz distribution is indicated in Fig. 6. In Fig. 5, the left ones are the results of two-step model, while the right ones are those of one-step model. As shown in Fig. 5, the number of outliers for two-step model is less than that for one-step model, especially in the range of high redshift. From the evaluation metrics of regression in Table 9 and outliers in Fig. 5, it is obvious that the performance of two-step model is better than that of one-step model. Fig. 6 further proves this result.

5.4 Application

Based on the above experimental results for the known samples, we put forward a workflow of photometric redshift estimation as shown in Fig. 7, which is applied to predict photometric redshifts of quasar candidates from BASS DR3. In Fig. 7, the red rectangle boxes indicate data analysis and black parallelograms represent intermediate data or results. The workflow includes six regressors

Table 8. The performance of photometric redshift estimation on different subsamples with different methods.

Sample	Input pattern	Method	Model parameter	MSE	MAE	Bias	σ_{NMAD}	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)	Time (s)
BSW_H	Pattern I	XGBoost	max_depth = 6 n_estimators = 500	0.0989	0.1515	−0.0254	**0.0253**	0.3144	0.4777	99.11	2.56	2
BSW_H	Pattern II	CatBoost	depth = 7 iterations = 3000	**0.0756**	**0.1439**	−0.0055	0.0256	**0.2750**	**0.5910**	**99.72**	**2.17**	5
BSW_H	Pattern III	RF	max_depth = 11 n_estimators = 1000	0.0781	0.1442	0.0035	**0.0253**	0.2796	0.5804	99.55	2.4	14
BSW_L	Pattern I	XGBoost	max_depth = 10 n_estimators = 700	0.1526	0.2646	−0.0017	0.1012	0.3907	0.6408	93.77	20.64	32
BSW_L	Pattern II	CatBoost	depth = 13 iterations = 3000	**0.1497**	**0.2624**	−0.0005	**0.1002**	**0.3870**	**0.6475**	**93.85**	**20.37**	285
BSW_L	Pattern III	RF	max_depth = 15 n_estimators = 500	0.1536	0.2656	0.0010	0.1012	0.3919	0.6386	93.67	20.68	1300
BS_W_H	Pattern IV	XGBoost	max_depth = 5 n_estimators = 1000	0.0800	0.1429	−0.0068	0.0250	0.2828	0.5119	99.51	**2.24**	3
BS_W_H	Pattern V	CatBoost	depth = 6 iterations = 2000	**0.0759**	**0.1404**	−0.0081	0.0248	**0.2756**	**0.5407**	**99.64**	2.32	5
BS_W_H	Pattern VI	RF	max_depth = 11 n_estimators = 300	0.0769	0.1405	−0.0007	**0.0242**	0.2773	0.5285	99.51	2.59	5
BS_W_L	Pattern IV	XGBoost	max_depth = 10 n_estimators = 1000	0.1791	0.2947	−0.0002	0.1109	0.4233	0.5967	92.72	23.66	370
BS_W_L	Pattern V	CatBoost	depth = 13 iterations = 3000	**0.1775**	**0.2934**	−0.0001	**0.1105**	**0.4213**	**0.6005**	**92.80**	**23.52**	450
BS_W_L	Pattern VI	RF	max_depth = 15 n_estimators = 1000	0.1812	0.2963	0.0006	0.1115	0.4257	0.5919	92.60	23.76	2811

Table 9. Comparison of the performance of photometric redshift estimation by two-step model with that by one-step model.

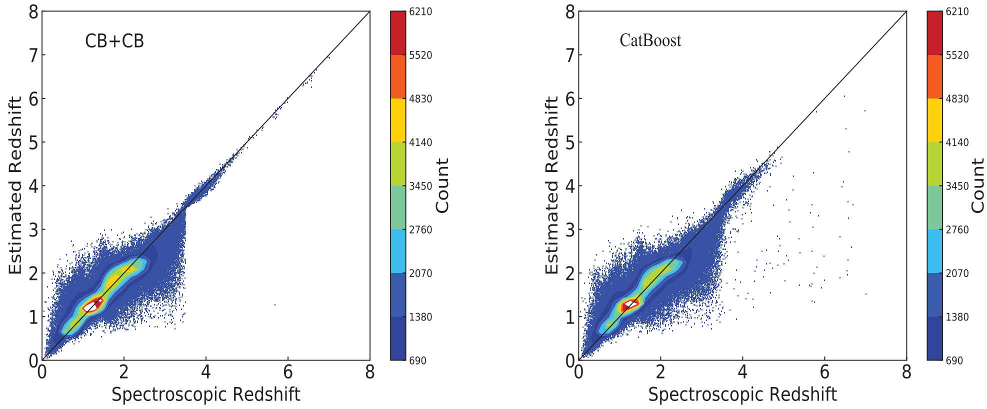
Test sample	MSE	MAE	Bias	σ_{NMAD}	$\sigma_{\Delta z}$	R^2	$\delta_{0.3}$ (per cent)	O (per cent)	Time (s)
			Two-step model						
BSW	0.0970	0.2153	−0.000 04	0.0854	0.3114	0.7967	96.32	15.13	25.0
BLW	0.1216	0.2103	0.0249	0.0784	0.3487	0.7630	94.86	14.81	2.0
BS_W	0.1266	0.2499	−0.000 04	0.0955	0.3558	0.7440	96.32	18.67	35.0
BL_W	0.1251	0.2157	0.0280	0.0804	0.3537	0.7553	94.40	15.63	9.0
			One-step model						
BSW	0.1059	0.2223	−0.000 02	0.0872	0.3254	0.7780	96.01	15.80	0.80
BLW	0.1239	0.2134	0.0265	0.0797	0.3521	0.7585	94.50	15.11	0.03
BS_W	0.1365	0.2578	−0.000 007	0.0981	0.3695	0.7239	94.84	19.42	1.00
BL_W	0.1277	0.2188	0.0296	0.0817	0.3574	0.7502	94.40	15.88	0.05

and one classifier. The detailed model information of this workflow is shown in Table 10.

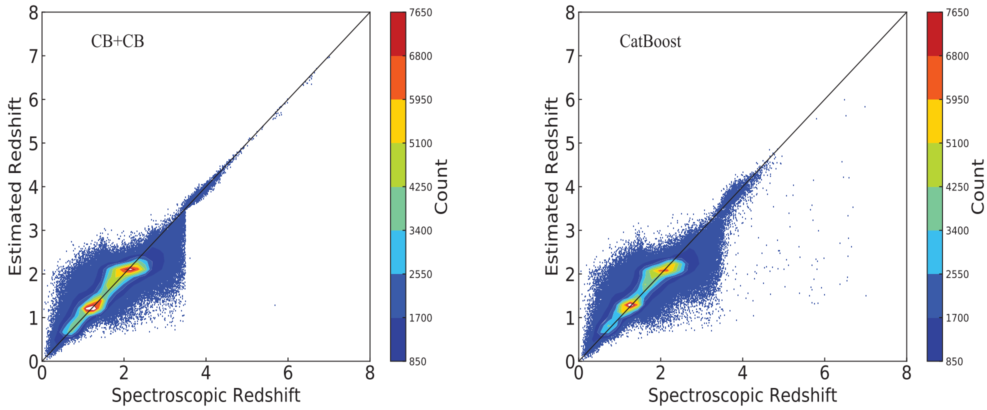
According to the work of Li et al. (2021), the BASS DR3 sources were identified as stars, galaxies, and quasars by XGBoost. We consider all possible quasar candidates, the total number of which is 26 200 778. We adopt the workflow to estimate their photometric redshifts.

In Fig. 7, BASS DR3 quasar candidates are first divided into two samples. One sample contains only optical features and the other sample contains optical and infrared features. For the candidates with optical and infrared features, we get predicted redshifts *redshift_p* by Regressor 1st, and these candidates are also classified into high redshift (high-Z) and low redshift (low-Z) subsamples. For the candidates classified as high-Z sources, we use Regressor 3rd to estimate redshifts, while for candidates classified as low-Z sources, Regressor 4th is used. Similarly, we obtain estimated redshifts of the candidates with only optical features. *redshift_bp* is predicted by two-step model while *redshift_p* is from one-step model. In the end, all predicted results are combined in a whole table. The link address is http://paperdata.china-vo.org/Li.Changhua/bass/bassdr3-quasar-z.hdf5. Table 11 lists 20 rows of predicted results, which is of great value for the further research on the characteristics and physics of these quasar candidates.

In the work of Li et al. (2021), when applying both optical and infrared information with the same predicted results by binary and multiclass classifiers, the number of quasar candidates is 2195 180, 1500 099 ($P_{\mathrm{Q}} > 0.75$), and 798 928 ($P_{\mathrm{Q}} > 0.95$). In Fig. 8, we show the number density distribution of quasar candidates as a function of photometric redshifts estimated by one- and two-step models. It is found from Fig. 8 that there is some difference between one- and two-step models for the predicted redshifts of the quasar candidates. Table 12 lists the number of BASS DR3 quasar candidates with $P_{\mathrm{Q}} > 0.95$ in different redshift ranges by the two models, which suggests that two-step



(a) For the sample BSW.



(b) For the sample BS_W.

Figure 5. Comparison of the photometric redshift with the spectroscopic redshift for the samples BSW and BS_W with two-step models (left-hand panel) and one-step models (right-hand panel), respectively.

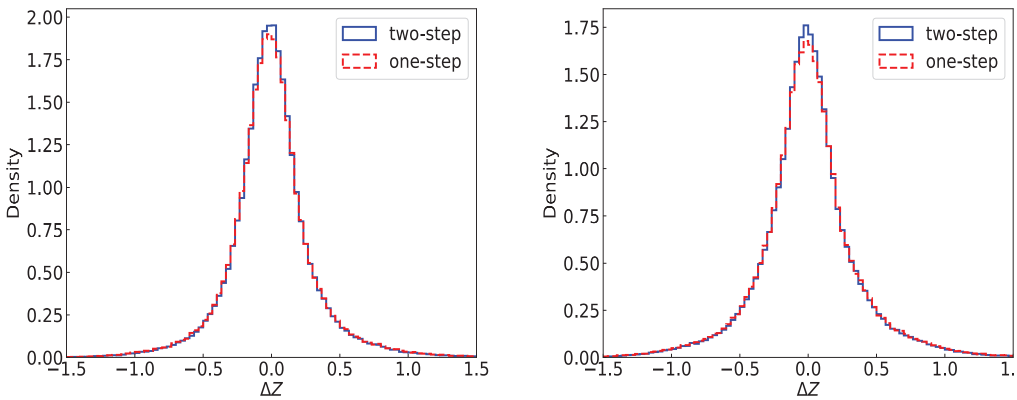


Figure 6. The Δz distribution for the samples BSW (left-hand panel) and BS_W (right-hand panel) with two-step models and one-step models during the range from -1.5 to 1.5, respectively.

model is more likely to find high-redshift quasars than one-step model.

6 CONCLUSIONS

We present two schemes of machine learning to predict photometric redshifts of quasars. We discuss the feature importance of different samples by XGBoost, CatBoost, and RF. The optimal features and optimal model parameters of these three algorithms are chosen for different samples. A comparison of classification and regression of these three algorithms is performed. By contrast, CatBoost achieves the best performance among all classifiers and regressors for different samples considering effectiveness and efficiency. Moreover, comparing the performance of two-step model with that of one-step

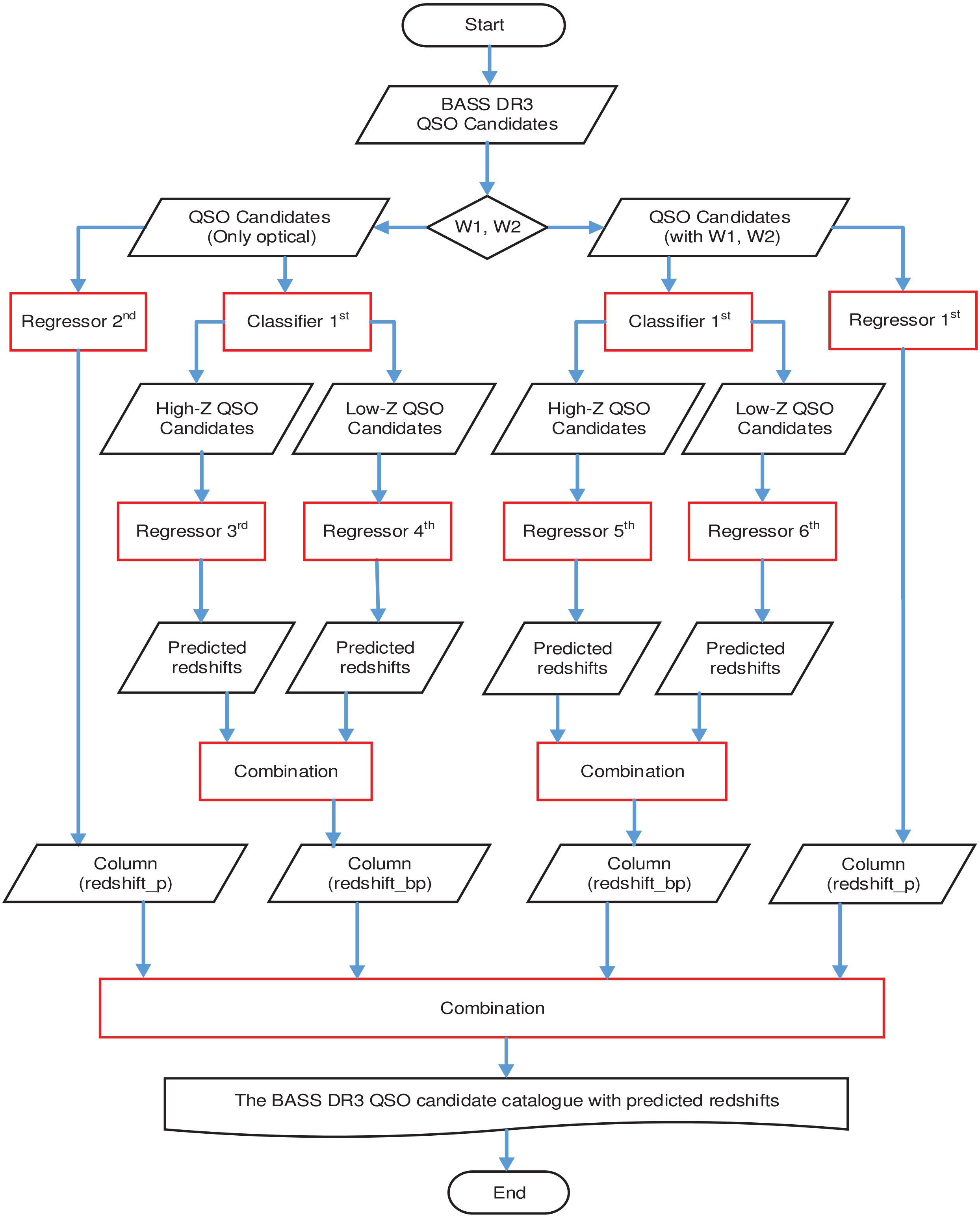


Figure 7. The photometric redshift estimation workflow.

model, two-step model is superior to one-step model. The two-step model is vital to select out high-redshift quasar candidates. Therefore, we put forward a photometric redshift workflow, in which CatBoost is taken as the core algorithm for classification and regression and two models are adopted. Then, we utilize the workflow to predict photometric redshifts of all quasar candidates from BASS DR3. The predicted result will be of great help and reference for future research of quasars and can help LAMOST, DESI, or other projects for follow-up observation to find more high-redshift quasars.

Table 10. The models used in photometric redshift estimation workflow.

Scheme	Method	Input pattern	Redshift range	$W1$, $W2$
Regressor 1st	CatBoost	Pattern II	Full	Indispensable
Regressor 2nd	CatBoost	Pattern V	Full	Dispensable
Regressor 3rd	CatBoost	Pattern II	Redshift $\geq$ 3.5	Indispensable
Regressor 4th	CatBoost	Pattern II	Redshift $<$ 3.5	Indispensable
Regressor 5th	CatBoost	Pattern V	Redshift $\geq$ 3.5	Dispensable
Regressor 6th	CatBoost	Pattern V	Redshift $<$ 3.5	Dispensable
Classifier 1st	CatBoost	Pattern III	(High redshift and low redshift)	Dispensable

Table 11. The estimated redshifts of BASS DR3 quasar candidates: redshift_*bp* is predicted redshift by two-step model, while redshift_*p* is predicted redshift by one-step model.

ID	RA	Dec.	redshift_*bp*	redshift_*p*
95429001151	133.484 498 720 996 38	84.647 600 582 453 69	3.612	2.755
95375007162	146.344 688 631 591 54	84.195 210 723 498 99	3.756	2.967
95375009802	146.706 830 963 958 08	84.340 119 918 879 81	3.853	3.071
95376013790	156.820 483 00 828 84	84.559 939 696 867 26	3.820	3.462
95432000953	156.684 107 846 550 92	84.621 645 961 979 78	3.770	2.216
95433002683	162.412 933 199 648 82	84.756 611 542 354 15	3.831	3.252
95379009607	172.106 422 949 517 34	84.540 703 842 548 78	4.290	2.970
95435000350	172.160 263 296 861 27	84.590 698 878 924 09	3.584	2.644
95439001178	204.459 789 679 880 97	84.631 479 439 353 14	3.703	2.324
95440003112	208.242 997 146 035 15	84.745 093 030 197 37	3.835	2.349
95440003256	207.357 202 487 6584	84.751 619 707 238 23	3.657	3.329
95441002154	214.027 878 464 509 85	84.718 400 042 0548	3.643	2.777
95372005855	128.463 761 795 6736	84.176 609 650 132 85	3.690	2.329
95372007632	127.969 383 336 721 99	84.254 551 133 126 14	3.818	3.120
95372008421	129.977 422 608 1903	84.301 344 013 783 02	3.753	2.407
95373007697	136.732 037 145 507 77	84.231 135 442 970 41	3.875	2.745
95373009831	138.717 829 675 324 23	84.324 912 026 284 85	3.725	2.288
95374010575	143.502 346 045 487 34	84.370 992 677 526 63	3.858	3.136
95312008225	151.103 937 733 8386	83.593 691 053 9942	3.716	3.013
95312012685	150.532 988 018 0199	83.788 115 730 972 89	3.861	1.869
95376009700	153.099 938 069 5831	84.324 668 989 6127	3.738	2.487

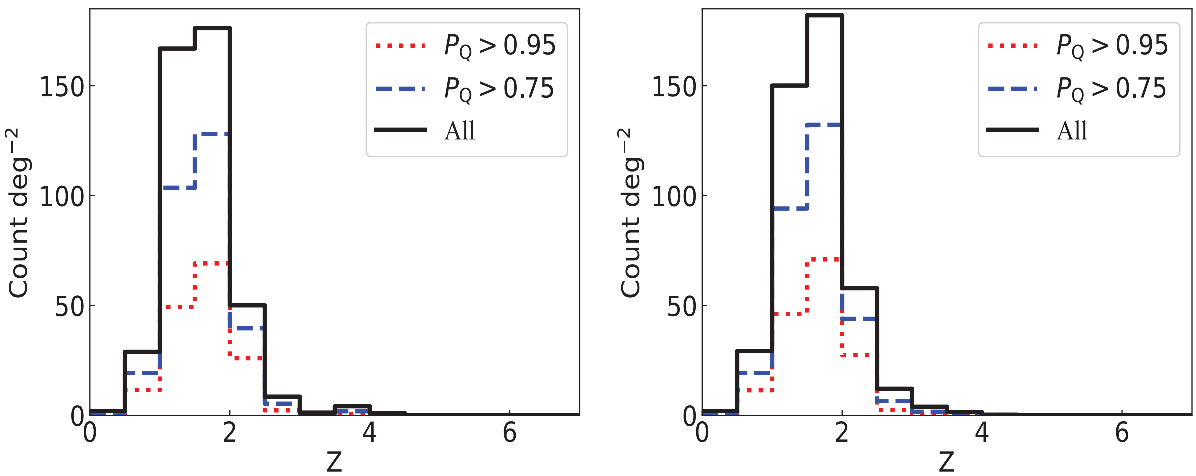


Figure 8. Left-hand panel: the number density of quasar candidates as a function of photometric redshifts by two-step model; right-hand panel: the number density of quasar candidates as a function of photometric redshifts by one-step model (Regressor 1st). In both panels, the quasar candidates contain both optical and infrared information with the same predicted results by binary and multiclass classifiers for different probabilities ($P_{\rm Q}$ > 0.95: red dotted line, $P_{\rm Q}$ > 0.75: blue dotted dash line, all: black line).

Table 12. The number of BASS DR3 quasar candidates with $P_{\rm Q} > 0.95$ in different redshift ranges by two models.

Model	Redshift <3.5	3.5 ≤ Redshift <4.5	4.5 ≤ Redshift <5.5	Redshift ≥5.5
One-step	796 078	2822	27	1
Two-step	794 990	3817	97	24

ACKNOWLEDGEMENTS

We are very grateful to the referee for his constructive suggestions and comments. This work is supported by National Natural Science Foundation of China (NSFC; grant numbers 11573019, 11803055, 11873066, 12133001, and 11433005), the Joint Research Fund in Astronomy (U1531246, U1731125, U1731243, and U1731109) under cooperative agreement between the NSFC and Chinese Academy of Sciences (CAS), the 13th Five-year Informatization Plan of Chinese Academy of Sciences (number XXH13503-03-107), and the science research grants from the China Manned Space Project with number CMS-CSST-2021-A06. We would like to thank the National R&D Infrastructure and Facility Development Program of China, 'Earth System Science Data Sharing Platform', and 'Fundamental Science Data Sharing Platform' (DKA2017-12-02-07). Data resources are supported by Chinese Astronomical Data Center (NADC) and Chinese Virtual Observatory (China-VO). This work is supported by Astronomical Big Data Joint Research Center, co-founded by National Astronomical Observatories, Chinese Academy of Sciences, and Alibaba Cloud. This research has made use of BASS DR3 catalogue. BASS is a collaborative program between the National Astronomical Observatories of the Chinese Academy of Science and Steward Observatory of the University of Arizona. It is a key project of the Telescope Access Program (TAP), which has been funded by the National Astronomical Observatories of China, the Chinese Academy of Sciences (the Strategic Priority Research Program. The Emergence of Cosmological Structures grant no. XDB09000000), and the Special Fund for Astronomy from the Ministry of Finance. BASS is also supported by the External Cooperation Program of the Chinese Academy of Sciences (grant number 114A11KYSB20160057). The BASS data release is based on the China-VO. The Guoshoujing Telescope (the LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

We acknowledge SDSS data bases. Funding for the SDSS-IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard–Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

DATA AVAILABILITY

The predicted photometric redshifts for BASS DR3 quasar candidates are saved in a repository and can be obtained by a unique identifier, part of which is indicated in Table 11. It is put in paperdata at http://paperdata.china-vo.org, and can be downloaded from the web link http://paperdata.china-vo.org/Li.Changhua/bass/bassdr3-quasar-z.hdf5.

REFERENCES

Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tcheng D., Llorà X., 2007, ApJ, 663, 774
Baum W. A., 1957, AJ, 62, 6
Blanton M. R. et al., 2017, AJ, 154, 28
Bonfield D. G., Sun Y., Davey N., Jarvis M. J., Abdalla F. B., Banerji M., Adams R. G., 2010, MNRAS, 405, 987
Breiman L., 2001, Mach. Learn., 45, 5
Brescia M., Cavuoti S., D'Abrusco R., Longo G., Mercurio A., 2013, ApJ, 772, 12
Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, ApJ, 712, 511
Carrasco K. M., Brunner R. J., 2014, MNRAS, 438, 3409
Cavuoti S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017, MNRAS, 465, 1959
Cavuoti S., Brescia M., Longo G., Mercurio A., 2012, A&A, 546, 8
Chen T., Guestrin C., 2016, Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Min. ACM, New York, NY
Cui X.-Q. et al., 2012, Res. Astron. Astrophys., 12, 1197
Curran S. J., Moss J. P., Perrott Y. C., 2021, MNRAS, 503, 2639
DESI Collaboration et al., 2016, preprint (arXiv:1611.00036)
Dorogush A. V., Ershov V., Yandex A. G., 2018, preprint (arXiv:1810.11363)
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Friedman J. H., 2001, Ann. Stat., 29, 1189
Han B., Qiao L.-N., Chen J.-L., Zhang X.-D., Zhang Y.-X., Zhao Y.-H., 2021, Res. Astron. Astrophys., 219, 17
Henghes B., Pettitt C., Thiyagalingam J., Hey T., Lahav O., 2021, MNRAS, 505, 4847
Hoyle B., 2016, Astron. Comput., 16, 34
Jin X., Zhang Y., Zhang J., Zhao Y., Wu X., Fan D., 2019, MNRAS, 485, 4539
Jones E., Singal J., 2017, A&A, 600, 11
Koo D. C., 1985, AJ, 90, 418
Leistedt B., Hogg D. W., 2017, ApJ, 838, 14

Li C. et al., 2017, in Brescia M., Djorgovski S. G., Feigelson Eric D., Longo G., Cavuoti S., eds, Proc. IAU Symp. 325. Cambridge University Press, Cambridge, p. 353
Li C. et al., 2021, MNRAS, 506, 1651
Luo A. L. et al., 2015, Res. Astron. Astrophys., 15, 1095
Mirabal N., Charles E., Ferrara E. C., Gonthier P. L., Harding A. K., Sanchez-Conde M. A., Thompson D. J., 2016, ApJ, 825, 69
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Schindler J., Fan X., McGreer I. D., Yang Q., Wu J., Jiang L., Green R., 2017, ApJ, 851, 13
Silva D. R. et al., 2016, Am. Astron. Soc. Meeting Abstr., 228, 317.02
Wang D., Zhang Y., Liu C., Zhao Y., 2007, MNRAS, 382, 1601
Way M. J., Foster L. V., Gazis P. R., Srivastava A. N., 2009, ApJ, 706, 623
Way M. J., Klose C. D., 2012, PASP, 124, 274
Way M. J., Srivastava A. N., 2006, ApJ, 647, 102
Wright E. L. et al., 2010, AJ, 140, 1868
Yèche Ch. et al., 2010, A&A, 523, A14
York D. G. et al., 2000, AJ, 120, 1579
Zhang Y.-X., Zhang J.-Y., Jin X., Zhao Y.-H., 2019, Res. Astron. Astrophys., 19, 169
Zhang Y., Li L., Zhao Y., 2009, MNRAS, 392, 233
Zhang Y., Ma H., Peng N., Zhao Y., Wu X.-b., 2013, AJ, 146, 10
Zou H. et al., 2017a, PASP, 129, 064101
Zou H. et al., 2017b, AJ, 153, 276
Zou H. et al., 2019, ApJS, 245, 4

This paper has been typeset from a TEX/LATEX file prepared by the author.