PhotoRedshift-MML: a multimodal machine learning method for estimating photometric redshifts of quasars

Shuxin Hong,$^{1,2,3,4}$ Zhiqiang Zou,$^{1,2}$ * A-Li Luo,$^{3,4}$ † Xiao Kong,$^3$ Wenyu Yang$^{1,2}$ and Yanli Chen$^{1,2}$

$^1$ College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
$^2$ Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing 210023, China
$^3$ CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Beijing 100101, China
$^4$ School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT

We propose a Multimodal Machine Learning method for estimating the Photometric Redshifts of quasars (PhotoRedshift-MML for short), which has long been the subject of many investigations. Our method includes two main models, i.e. the feature transformation model by multimodal representation learning, and the photometric redshift estimation model by multimodal transfer learning. The prediction accuracy of the photometric redshift was significantly improved owing to the large amount of information offered by the generated spectral features learned from photometric data via the MML. A total of 415,930 quasars from Sloan Digital Sky Survey (SDSS) Data Release 17, with redshifts between 1 and 5, were screened for our experiments. We used $|\Delta z| = \left(\frac{(z_{\text{phot}} - z_{\text{spec}})}{1 + z_{\text{spec}}}\right)$ to evaluate the redshift prediction and demonstrated a 4.04% increase in accuracy. With the help of the generated spectral features, the proportion of data with $|\Delta z| < 0.1$ can reach 84.45% of the total test samples, whereas it reaches 80.41% for single-modal photometric data. Moreover, the Root Mean Square (RMS) of $|\Delta z|$ is shown to decreases from 0.1332 to 0.1235. Our method has the potential to be generalized to other astronomical data analyses such as galaxy classification and redshift prediction.

Key words: methods: data analysis – techniques: photometric – surveys – methods: statistical

1 INTRODUCTION

Quasars are extremely bright active galactic nuclei (AGN), one of the brightest, most powerful and most energetic objects known in the universe (Peterson 1997), and a critical class of samples for studying cosmology and galaxy physics. With the Wide Field Survey, more than 750,000 quasars have been discovered (Lyke et al. 2020), most of them from the Sloan Digital Sky Survey (SDSS, York et al. 2000). According to Hubble’s law, quasars that are generally farther away tend to have larger redshifts. To determine the redshift accurately, a spectroscopic observation is required. However, the magnitude limit in the telescope’s observation spectrum is often lower than that of photometric observation, which necessitates the selection of a target. On the other hand, despite the existence of spectroscopic observations, not enough features are available in some particular wavelength’s band to enable an accurate redshift determination. Therefore, obtaining a complete spectroscopic redshift of quasars in a short time remains challenging in this case. In contrast, the photometric redshift ($z_{\text{phot}}$) is inferred from the multi-band photometric data set, which greatly reduces the resources budget and can provide a relatively complete data sample for the study of galaxy formation, evolution and cosmology. However, increasing the precision of the photometric redshift measurement constitutes an essential requirement for this sort of investigations.

Various techniques are available to obtain photometric redshifts, ranging from template-fitting to machine learning (hereafter ML) and hybrid systems (Salvato et al. 2019). Here, we focus on the latter two types of methods. At present, data mining and ML technologies are being widely used for dealing with massive astronomical spectroscopic and photometric data, owing to their high efficiency and accuracy (Brescia et al. 2021; Fluke & Jacobs 2020; Baron 2019). They can be divided into two categories: supervised and unsupervised learning. Supervised ML methods require training data with both photometric data as inputs and spectroscopic redshifts as labels. Gaussian Processes (GP) can be used to produce photo-z estimations (Duncan et al. 2018; Ansari et al. 2021). Tree-based algorithms, such as decision trees and random forests, are often utilized to provide an estimation of photometric redshifts (Carliles et al. 2010; Hoyle et al. 2015; Mountrichas et al. 2017). Support vector machine (SVM) (Peng et al. 2010) and K-Nearest Neighbour (KNN) (Ball et al. 2008; Curran 2020) are also widely used for their simplicity. Neural Networks are popular supervised ML algorithms (Bonnett 2015) including methods such as artificial neural networks (ANN) (Yèche et al. 2010; Sadeh et al. 2016), deep neural networks (DNN) (Hoyle 2016), convolutional neural networks (CNN) (D’Isanto & Polsterer 2018; Pasquet et al. 2019; Mu et al. 2020), and many others. The Multi-Layer Perceptron with Quasi-Newton Algorithm (MLPQNA) is another very popular deep learning method for estimating photo-z (Cavuoti et al. 2012, 2017; Brescia et al. 2019; Razim et al. 2021), among which Brescia et al. (2019) computed photometric redshift for AGN and compared it to spectral energy distribution (hereafter...
Multimodal machine learning (hereafter MML) (Baltrušaitis et al. 2018) has become a research hotspot in recent years for it is closer to the human learning style. Wu et al. (2014) fused image and audio data through multimodal representation learning and achieved a better classification effect compared to the case of single-modal data. Mroueh et al. (2015) fused the mouth image data and sound data of human speech based on the multimodal deep learning model, which improved the recognition effect of mouth images. Owens et al. (2016) learned the features of two modal data (video and audio) through CNN and Long Short Term Memory (LSTM) and successfully simulated the audio data from the video data. Transfer learning in multimodal machine learning is to use information from one resource-rich modality to assist another relatively resource-poor modality to learn, which can solve the problems of data dependence and so on. Fu et al. (2021) considered the inconsistent data distribution of quasars in low silver and high Galactic latitudes and successfully constructed the missing quasar optical data in low silver latitude using the existing quasar optical data in high silver latitude based on the transfer learning.

Inspired by the idea of supervised MML, we introduce MML to represent the photometric and spectroscopic data of SDSS formally through transferring photometric data to the feature space of spectra and using the converted features to assist the prediction of quasars’ photometric redshift. Specifically, the high-dimensional spectroscopic data are represented, and then the photometric magnitude is transferred to the spectral feature space. There are two kinds of data consisting of SDSS spectrum and photometry in the training samples, but only SDSS photometry data in the testing samples. In the pre-training stage, CNN with an attention mechanism is used to learn feature knowledge from high-dimensional spectroscopic data. Then, we build a generative model to train photometric data to generate simulated spectral data, which are very close to the true spectra, so as to assist the photometric data in the downstream task to complete the redshift prediction and improve its accuracy. To the best of our knowledge, few people have used the multimodal transfer learning method to predict the photometric redshift of quasars. The experimental results show that our method significantly improves the accuracy of photometric redshift estimation of quasars.

The main contributions of this study are as follows: (1) A photometric to spectral feature transformation model (i.e., PhotoSpecTransformer), based on multimodal representation learning, is constructed to represent data pairs (photometric data features, spectroscopic data features) in a unified feature space. The goal of the PhotoSpecTransformer model is to generate simulated spectral data from photometric data; (2) A photometric redshift estimation model based on transfer learning (i.e., PhotoReshift-TL) is constructed to assist the photometric data with the generated simulated spectral data, so as to improve the accuracy of redshift prediction; (3) Our method has generality to some extent. If the even pair data of quasars are replaced by that of galaxies, the redshift of galaxies can also be predicted with higher accuracy; (4) We share the codes and experimental data, published in GitHub (https://github.com/HongShuxin/PhotoRedshift-MML), to facilitate the access to their usage by other researchers to reproduce the research and further expand.

The remaining of this paper is organised as follows. In Section 2, we introduce the acquisition, composition, and pre-processing of data. In Section 3, we present the method used in this paper, its implementation process, and the structure of the models in detail. Section 4 gives the evaluation metrics, comparisons, and analyses of experimental results. Finally, Section 5 summarizes and discusses the key findings of our work.
In the following, we introduce the processing methods of photometric and spectroscopic data, respectively.

2.2.1 Pre-processing of photometric data

For photometric data, extinction correction should first be carried out in advance. The magnitude of each band should be subtracted from the extinction of the corresponding band, given by equation (1) as:

$$\text{mag}_k = \text{psfMag}_k - \text{extinction}_k,$$

where $k$ represents one of the five bands, $\text{psfMag}_k$, $\text{extinction}_k$, and $\text{mag}_k$ represent the original magnitude, the extinction, and the magnitude after extinction correction of $k$-band, respectively.

Secondly, to eliminate the influence of distance, and increase the number of features, we calculate the values of five colours, which are $\text{mag}_u - \text{mag}_g$, $\text{mag}_g - \text{mag}_r$, $\text{mag}_r - \text{mag}_i$, $\text{mag}_i - \text{mag}_z$, $\text{mag}_u - \text{mag}_z$, respectively. So far, we have got 15 dimensional photometric data, i.e., $\text{mag}_u$, $\text{mag}_g$, $\text{mag}_r$, $\text{mag}_i$, $\text{mag}_z$, $\text{mag}_u - \text{mag}_g$, $\text{mag}_g - \text{mag}_r$, $\text{mag}_r - \text{mag}_i$, $\text{mag}_i - \text{mag}_z$, $\text{mag}_u - \text{mag}_z$, $\text{psfMagErr}_u$, $\text{psfMagErr}_g$, $\text{psfMagErr}_r$, $\text{psfMagErr}_i$, $\text{psfMagErr}_z$, as the input features of the quasar redshift prediction model.

Thirdly, before inputting into the model, the 15 features should be normalized for the sake of measurement’s consistency. Considering that their numerical range is relatively concentrated, linear normalization is adopted. The calculation method is shown in equation (2). The calculation results would build a 15-dimensional vector, denoted as $\chi_{\text{phot}}$.

$$X_j\_\text{norm} = \frac{X_j - X_j\_\text{min}}{X_j\_\text{max} - X_j\_\text{min}},$$

where $X_j$ represents the original value of the $j$-th feature ($1 \leq j \leq 15$), $X_j\_\text{min}$ and $X_j\_\text{max}$ represent the minimum and maximum values of the $j$-th feature, respectively, and $X_j\_\text{norm}$ represents the $j$-th eigenvalue after normalization.

2.2.2 Pre-processing of spectroscopic data

For spectroscopic data, we should first ensure the consistency of input dimensions. Considering the inconsistency of skylight residues at the red end of spectroscopic data (Hong et al. 2021; Zou et al. 2020), the first 3600 dimensional features of spectroscopic data are uniformly retained. In this process, it is found that the dimension of a small number of spectra is less than 3600, so this part of spectra and their corresponding photometric data are removed from the data set. Further, there are 415,930 quasar data left from the original 416,296 quasars.

Secondly, spectroscopic data also need to be normalized. The flux at different wavelengths of the spectrum varies greatly. If not treated, the features of high intensity will dominate, whereas the low intensity features will be ignored by the model. To address this issue, we use the method called flux standardization (Li et al. 2007) to normalize the spectroscopic data, which could eliminate the influence of dimension and accelerate the convergence speed of the model. The specific calculation method can be seen in equation (3). The calculation results can form a 3600-dimensional vector, denoted as $\chi_{\text{spec}}$.

$$y_m = \frac{y_m}{\|y_m\|_2},$$

where $y_m$ represents the $m$-th spectrum, $\|y_m\|_2$ represents the 2-norm of the $m$-th data obtained by computing the square root of the sum of squares of each element.
3 METHODS

Considering that the quasar photometric data contains too few features, which leads to low photometric redshift prediction accuracy, a Photometric Redshift estimation based on the Multimodal Machine Learning (PhotoRedshift-MML) method is proposed to improve the accuracy of photometric redshift prediction, as shown in Figure 2. PhotoRedshift-MML consists of two parts, one is a photometric to spectral feature transformation model (PhotoSpecTransformer), whose specific implementation will be introduced in Section 3.2.1. The other is a photometric redshift estimation model (PhotoRedshift-TL), whose specific implementation will be introduced in Section 3.2.2. In order to verify the advantage of the multimodal method over the single-modal method, we first constructed a photometric redshift prediction model based on single-modal machine learning in Section 3.1, which is used as one of the baselines of our work.

3.1 Prediction of quasars’ photometric redshifts based on single-modal machine learning

With only 15 input features for single-modal photometric data, there is no need to use a too deep network or complex model. We constructed a four-layer ANN model to convert the input into the feature space, and then predict the photometric redshift of quasars. Figure 3 illustrates the model structure of the ANN constructed here.

The model is composed of one input layer, two hidden layers, and one output layer. Each circle in Figure 3 represents one neuron. The input layer is equipped with 15 neurons corresponding to the 15 features of the input. The first hidden layer is equipped with 32 neurons, the second hidden layer with 16 neurons, and the last layer outputs the predicted redshift value. Starting from the second layer, the input of neurons in each layer is a linear combination of neurons in the previous one. Each hidden layer is added with \( \text{relu} \) activation function (shown in equation (4)), providing nonlinear variation.

\[
\text{relu} = \max(0, x),
\]

where \( \max \) is the function of calculating the maximum value, and \( x \) represents the output of the previous layer network.

It can be seen that the structure of the above model is straightforward. We have also tried to increase the depth of the model or use other complex models, such as CNN. Nonetheless, the complex model would significantly increase the training time, without necessarily better prediction results than those of the above model. Having only 15 input features, an overly complex model would thus lead to overfitting. For this, we finally chose this simple model to get better results at the lowest cost.

In the process of model training, we define the loss function as \( L_1 \) (denoted as loss(\( L_1 \))). It represents the average absolute error amplitude of the predicted value, as shown in equation (5). Compared to the mean square error, loss(\( L_1 \)), chosen here, has better robustness to outliers.

\[
\text{loss}(L_1) = \frac{\sum_{n=1}^{N} |z_{\text{spec}}^n - z_{\text{phot}}^n|}{n},
\]

where \( N \) represents the total number of samples (i.e. 415,930), \( n \) ranges from 1 to \( N \), \( z_{\text{spec}}^n \) represents the spectroscopic redshift of the \( n \)-th data, and \( z_{\text{phot}}^n \) represents the predicted photometric redshift of the \( n \)-th data.

3.2 Design and implementation of the PhotoRedshift-MML method

In Section 3.1, the single-modal photometric data is used to predict photometric redshifts of quasars. As mentioned in the introduction, because of the insufficient features available for model learning, this single-modal machine learning model has encountered a bottleneck with difficulty to improve its prediction accuracy. Therefore, in Section 3.2.1, multimodal representation learning is introduced to represent data pairs (photometric data features, spectroscopic data features) in a unified feature space. The represented spectral features are used as labels to generate simulated spectral data. In 3.2.2, aiming at the problem of low redshift prediction accuracy due to few photometric features of quasars, the simulated spectra are used to assist the photometric redshift prediction of quasars. Thus, a photometric redshift prediction model based on multimodal transfer learning is constructed.

3.2.1 Feature transformation model based on multimodal representation learning

In some cases, only the photometric data of quasars exist due to the limitation of observation equipment. In this case, the prediction of
quasar redshift can only rely on a small number of features contained in photometric data, which renders it challenging to obtain high prediction accuracy. This section aims at considering the knowledge of the real spectra as labels and transfer photometric data into spectral features so that the prediction accuracy can still be improved with the help of simulated spectra using photometric data alone. To acquire knowledge from the pre-training and achieve the conversion from photometric data to spectral features, the following two steps are required. First, in a unified feature space, the data pairs (photometric data features, spectroscopic data features) are constructed, denoted as \( \mathcal{F}_0 \), and its input as \( \chi_{\text{spec}} \), and its output as \( \chi_{\text{spec}}^{\text{true}} \), referring to equation (6),

\[
\chi_{\text{spec}}^{\text{true}} = \mathcal{F}_0(\chi_{\text{spec}}; \theta_0),
\]

where \( \theta_0 \) is a vector composed of all the parameters in the network \( \mathcal{F}_0 \).

Then, data pairs (photometric data features, spectroscopic data features) are constructed, denoted as \( (\chi_{\text{phot}}; \chi_{\text{spec}}^{\text{true}}) \). Taking \( \chi_{\text{spec}}^{\text{true}} \) as labels, we build a PhotoSpecTransformer model, named \( \mathcal{F}_1 \), referring to equation (7). \( \chi_{\text{phot}} \) are used to generate simulated spectral features, denoted as \( \chi_{\text{spec}}^{\text{generated}} \).

\[
\chi_{\text{spec}}^{\text{generated}} = \mathcal{F}_1(\chi_{\text{phot}}; \chi_{\text{spec}}^{\text{true}}; \theta_1),
\]

where \( \theta_1 \) is a vector composed of all the parameters in the model \( \mathcal{F}_1 \).

So far, we have got simulated spectral features, \( \chi_{\text{spec}}^{\text{generated}} \), so that we can enrich our features to predict redshifts, and then we build the PhotoRedshift-TL model, named \( \mathcal{F}_2 \). Taking \( (\chi_{\text{phot}}; \chi_{\text{spec}}^{\text{generated}}) \) as input, we obtain the prediction results of photometric redshift, denoted as \( \chi_{\text{phot}} \), as shown in equation (8). The detail of \( \mathcal{F}_2 \) will be presented in Section 3.2.2.

\[
Z_{\text{phot}} = \mathcal{F}_2(\chi_{\text{phot}}; \chi_{\text{spec}}^{\text{generated}}; \theta_2),
\]

where \( \theta_2 \) is a vector composed of all the parameters in the model \( \mathcal{F}_2 \).

Briefly, our PhotoRedshift-TL method is \( \mathcal{F}_2(\mathcal{F}_1(\mathcal{F}_0)) \), as generalized in Figure 2. The model \( \mathcal{F}_0 \) is implemented based on a convolutional neural network, as shown in Figure 4. It has the pre-processed 3600-dimensional spectroscopic data as input, predicting photometric redshift as the final goal, and the \( \text{loss}(L1) \) function in equation (5) as the objective function. After training, we retain the 128-dimensional spectral vector obtained by the model \( \mathcal{F}_0 \) as the label in step (2).

The implementation of model \( \mathcal{F}_0 \) is described in the following five parts:

(i) One input layer;
(ii) regard four one-dimensional convolutional layers with attention blocks and one pooling layer as a module (denoted as Convolutional-Attention-Pooling, CAP module for short). There are four modules with the same structure but different filter numbers (the number of filters in convolutional layers of each module is 128, 64, 32, and 16, respectively. Therefore, they are denoted as CAP128, CAP64, CAP32 and CAP16);
(iii) one flatten layer;
(iv) nine fully connected layers (the number of neurons is reduced by half from 2048 to 8 gradually, and dropout function is introduced to randomly discard half of the neurons with a probability of 0.5 to prevent overfitting);
(v) one output layer.

(2) Step 2: PhotoSpecTransformer model construction

The PhotoSpecTransformer model, i.e. \( \mathcal{F}_1 \) model, is based on ANN. As shown in Figure 5, the input of the model are the features of 15-dimensional photometric data, \( \chi_{\text{phot}} \), and the 128-dimensional spectral feature \( \chi_{\text{spec}}^{\text{true}} \) obtained in step (1), whereas the output of the model is \( \chi_{\text{spec}}^{\text{generated}} \). That is, we can generate the 128-dimensional spectral feature from the 15-dimensional photometric data using the \( \mathcal{F}_1 \) model.

The implementation of model \( \mathcal{F}_1 \) is described as follows:

(i) one input layer (including 15 neurons corresponding to the 15-dimensional input photometric data);
(ii) four hidden layers (containing 8, 16, 32 and 64 neurons, respectively). Each hidden layer is also added with \( \text{relu} \) activation function (see equation (4)), providing nonlinear variation;
(iii) one output layer.

In the process of model \( \mathcal{F}_1 \) training, we iteratively optimize the cosine loss function, shown in equation (9).

\[
\log \text{Cosh} = \sum_{n=1}^{N} \log(\cosh(\chi_{\text{spec}}^{\text{generated}}, \chi_{\text{spec}}^{\text{true}})),
\]
and a testing set in the proportion of 8:2. The experimental results

Figure 6. The structure of model $\mathcal{F}_2$.

where $\chi_{\text{spec}}^{\text{true},n}$ represents the 128-dimensional true feature vector of the $n$-th spectroscopic data, $\chi_{\text{spec}}^{\text{generated},n}$ represents the 128-dimensional simulated spectral vector.

So far, we have successfully built the PhotoSpecTransformer model, which we can further train and use. In the training stage of the model, the photometry-spectra pairs are used to iteratively optimize the parameters of the model by continuously narrowing the gap between the generated and real spectra, so that the model can generate enough similar spectral features. In the application stage of the model, the simulated spectral features can be generated adopting photometric data alone, which can assist the downstream task, i.e. quasar photometric redshift prediction.

3.2.2 Prediction of quasar photometric redshifts based on multimodal transfer learning

In Section 3.2.1, the 128-dimensional simulated spectral vector $\chi_{\text{spec}}^{\text{generated},n}$, generated from the photometric data, has been obtained. Next, these simulated spectral vectors can be used to assist the 15-dimensional photometric data, $\chi_{\text{phot}}$, to more accurately predict the photometric redshift of quasars.

The PhotoRedshift-TL model, i.e. model $\mathcal{F}_2$, similar to model $\mathcal{F}_0$, is built on CNN and CAP module, as shown in Figure 6. The specific structure will not be discussed in detail here, we only discuss the differences, which lie in the input layer and the fully connected layers.

(i) The input of model $\mathcal{F}_2$ has two parts: one is Main_input, which consists of the features of 15-dimensional photometric data $\chi_{\text{phot}}$, the other is Auxiliary_input, which is the 128-dimensional simulated spectral features $\chi_{\text{spec}}^{\text{generated}}$ obtained in 3.2.1. The two parts of data are concatenated to form a new feature vector $V$ of 143 * 1, which is fed into CAP64 and CAP16.

(ii) The model $\mathcal{F}_2$ has three fully connected layers (containing 64, 32 and 16 neurons, respectively). Since not many neurons are available in model $\mathcal{F}_2$, the dropout function is not used to discard neurons.

4 RESULTS AND DISCUSSION

In order to verify the performance of PhotoRedshift-MML, this section uses the pre-processed data in Section 2, i.e. $(\chi_{\text{phot}}, \chi_{\text{spec}})$. In the experiment, 415,930 pairs of data are divided into a training set and a testing set in the proportion of 8:2. The experimental results are evaluated by three evaluation metrics defined in Section 4.1, and the results are compared and analysed with baselines in Section 4.2.

4.1 Evaluation metrics

We use six metrics to verify the prediction performance of the model. For the prediction of galaxy redshift, the commonly used statistical mean square error, i.e. MSE, is selected as the evaluation metric. The calculation method is shown in equation (10). For the prediction of quasar redshift (Zhang et al. 2013), we introduce classical $|\Delta z|$ and Root Mean Square (RMS) of $|\Delta z|$, defined in equation (11) and equation (12), respectively. To further measure the quality from different perspectives, we also calculate the fraction of outliers (Cavuoti et al. 2012), defined in equation (13), normalized median absolute deviation (NMAD) (Ilbert et al. 2008), defined in equation (14) and bias (Cavuoti et al. 2012), defined as mean($\Delta z$) after excluding outliers.

\[
\text{MSE} = \frac{\sum_{n=1}^{N}(z_{\text{spec}}^{n} - z_{\text{phot}}^{n})^2}{N},
\]

\[
|\Delta z| = \left| \frac{z_{\text{phot}}^{n} - z_{\text{spec}}^{n}}{1 + z_{\text{spec}}^{n}} \right|,
\]

\[
\text{RMS} = \sqrt{\frac{\sum_{n=1}^{N}|\Delta z|^{2}}{N}}.
\]

\[
|\Delta z| > 0.15,
\]

\[
\sigma_{\text{NMAD}} = 1.48 \times \text{median}(|\Delta z|),
\]

where $z_{\text{spec}}^{n}$ represents the spectroscopic redshift of the $n$-th data, $z_{\text{phot}}^{n}$ represents the predicted photometric redshift of the $n$-th data, $|\Delta z|_{n}$ represents the $|\Delta z|$ of the $n$-th data.

4.2 Comparison and analysis

The experimental environment of this paper is as follows. We use an Intel Xeon E5-2690v2 2.6GHz CPU and an Nvidia Tesla K40 12GB GPU. The software environment includes Python 3.5, Keras 2.3.1, Numpy 1.16.2, Matplotlib 3.0.3 and Scikit-learn 0.19.1.

In terms of parameters selection, the batch_size is set to 512, the optimizer selects Adam, and the learning rate is 0.001.

(1) Experiment 1: verifying the generality

The generality of the method is reflected in that our model can also be used to perform other downstream tasks, such as predicting the photometric redshifts of galaxies with high accuracy, either using the model presented in Section 3.1 or Section 3.2. Considering that the MSE of galaxy photometric redshift prediction obtained by the model described in Section 3.1 is already low, which reaches 0.0011, there is no need to use multimodal machine learning to increase accuracy. Taking Mu et al. (2020) as the Baseline 1, using the same galaxy data set, we use ANN proposed in Section 3.1 instead of CNN to predict the photometric redshift of galaxy data and calculate the MSE of the prediction results. It is found that the prediction accuracy is equivalent (even slightly better) to that in Baseline 1, as shown in Table 1. This is because when there are few input characteristics of the model, their CNN is not the optimal choice. See Table 1 for the specific comparison results. As for the generality of
A multimodal machine learning for Photo-z

Figure 7. Prediction results of photometric redshift based on ANN.

Figure 8. Prediction results of photometric redshift based on PhotoRedshift-MML.

Table 1. Comparison of Galaxy photometric redshift prediction results between Baseline 1 and our method.

| Data Set                        | MSE    |
|---------------------------------|--------|
| Results of Baseline 1           |        |
| SDSS DR13(early-type galaxy)    | 0.0014 |
| SDSS DR13(late-type galaxy)     | 0.0019 |
| Our results                     |        |
| SDSS DR13(galaxy)               | 0.0011 |

The abscissa of Figure 7 (a) and Figure 8 (a) is $z_{\text{spec}}$ with the ordinate being $z_{\text{phot}}$. The prediction results are drawn on the left part of the figures, and the corresponding colour bar is described on the right part of figures. As for the colour bar, the more points, the redder the colour, whereas for less points, the bluer is the colour. Figure 7 (b) and figure 8 (b) show the statistical number of samples with $z_{\text{phot}}-z_{\text{spec}}$ in different intervals. The abscissa is $z_{\text{phot}}-z_{\text{spec}}$, the bin is 0.1, the middle bin ranges from -0.05 to +0.05, and the ordinate is the number of samples. Comparing Figure 7 (a) to Figure 8 (a) shows that the error of redshift prediction based on PhotoRedshift-MML is significantly reduced and the data is more concentrated near the diagonal, indicating that the predicted value is closer to the real one. In other words, it is effective to add the simulated spectra obtained through transfer learning as an auxiliary input. For example, the proportion of data with $|\Delta z|$ within 0.1 increases from 80.41% to 84.45%. Comparing Figure 7 (b) to Figure 8 (b) shows that the amount of data with a smaller prediction error of ±5% has increased by a quarter from 17502 to 21780. In comparison, the amount of data with a larger prediction error has been significantly reduced. There are two reasons for the above results. One is that our PhotoRedshift-MML method can make good use of photometric data and simulated...
spectral vectors. The other is that the number of features increases from 15 to 143 by adding the simulated spectral vector close to the real spectral vector.

We calculate the percentages of different $|\Delta z|$ intervals and RMS of $|\Delta z|$ of the above experimental results, by taking Zhang et al. (2013) as the Baseline 2. The results are shown in Table 2.

Zhang et al. (2013) uses different datasets from SDSS DR7 and SDSS DR7-WISE. Their experimental results show that the more data sets used, the richer the information contained and the higher the accuracy of predicting photometric redshift. When using SDSS data alone, the proportion of data with $|\Delta z| < 0.1$ accounts for 77.13% and 80.41%, respectively, which are similar to 78.63% in Baseline 2. However, the proportion of data with $|\Delta z| < 0.2$ accounts for 94.47% and 92.60%, respectively, which are much higher than that of Baseline 2. Moreover, the proportion of data with $|\Delta z| < 0.3$ accounts for 97.13% and 96.06%, respectively, which are much higher than that of Baseline 2. Our RMS of $|\Delta z|$ is 0.1174 on SDSS DR8 and 0.1332 on SDSS DR17, whereas that in Baseline 2 is 0.259. This indicates that the overall performance of our ANN model is much better than Baseline 2. After adding the generated simulated spectral vector as an auxiliary input on SDSS DR8, the proportion of data with $|\Delta z|$ within 0.1, obtained by our PhotoRedshift-MML method, increases to 90.78%, which is about 13% higher than that of a single modal method. Whereas on SDSS DR17, the proportion increases by 4.04% (i.e. proportion of data reaches 84.45%). The proportion of data with $|\Delta z| < 0.2$ and $|\Delta z| < 0.3$ also increase, although they are not very significant because they are already very high in single modal data. RMS of $|\Delta z|$ decreases from 0.1174 to 0.1046 on SDSS DR8 and from 0.1332 to 0.1235 on SDSS DR17, which shows that the photometric redshift predicted by multimodal transfer learning is more accurate than that predicted by using photometry alone on the whole. Throughout the four statistical results of PhotoRedshift-MML, the effect of using photometric data fused with 128-dimensional simulated spectral features from $\mathcal{F}_1$ model can be comparable to that of using SDSS-WISE.

(3) Quality assessment and comparisons.

In order to evaluate the prediction results from different perspectives, we also calculate three commonly used statistical metrics: the fraction of outliers, NMAD, and bias. The results can be seen in Table 3.

Typically, outliers are those objects who meet equation (13), which is a quality control metric (Cunha & Humphrey 2022). On SDSS DR17, with single modal method, outliers account for 10.50%, whereas with PhotoRedshift-MML method, outliers account for 9.05%. The outlier fraction reduces by 1.45% with the auxiliary of generated spectral features.

The NMAD, defined in equation (14), provides a measure of the variability in the sample. This dispersion estimate is less sensitive to outliers (Brammer et al. 2008). The results of $\sigma_{NMAD}$ are 0.0732 and 0.0520 for single modal method and multi-modal method, respectively.

The bias shows the average separation between prediction and true values (Li et al. 2022). It is quite small in both methods. Particularly, the bias in PhotoRedshift-MML method is only -0.0042, which is 0.015 closer to zero than the single modal method. The closer the bias is to zero, the smaller the systematic bias is, between photometric and spectroscopic redshifts (Dahlen et al. 2013).

5 CONCLUSION

Although machine learning is widely used in estimating quasar photometric redshifts, we believe that the present work provides a first development of a quasar photometric redshift prediction method based on multimodal machine learning in this field. Furthermore, we demonstrate the effectiveness and advantage of our PhotoRedshift-MML by conducting extensive experiments. Specifically, in the pre-training stage, the photometric data are used to learn the spectral feature representation, and then the photometry-spectra feature transformation model is constructed. The trained model can make the photometric data generate very similar spectral feature vectors. In the downstream task, only photometric data is available, the prediction accuracy of photometric redshift can be significantly improved with the assistance of simulated spectral feature vectors.

In order to verify the effectiveness, advantage and generality of our method, we used the data from SDSS DR8 and DR17 to form an even pair (photometric data, spectroscopic data) for the same quasar observation source. The major advantage of PhotoRedshift-MML is that we first learn knowledge of the real spectra, and then convert photometric data to spectral features by transfer learning. This greatly reduces the reliance of collecting spectroscopic data to get more input features.

In terms of effectiveness of the method, we used SDSS DR8 and DR17 photometric data alone for single-modal redshift prediction. Due to the fact that the input features are less and small, we build a 4-layer ANN network. The predicted quasar photometric redshift of $|\Delta z| < 0.1$ accounts for 77.13% and 80.41% of the total test samples from SDSS DR8 and DR17 respectively, which are close to 78.63% obtained by Zhang et al. (2013). However, the proportion of data with $|\Delta z| < 0.2$ and $|\Delta z| < 0.3$, both vastly increased compared to that in Zhang et al. (2013). The RMS of $|\Delta z|$ has also decreased greatly. Therefore, our method has certain effectiveness.

In terms of advantage of the method, we used multimodal data to predict the redshift. After adding the additional features of the simulated spectrum, the proportion of data with $|\Delta z| < 0.1$ accounts for 90.78% on SDSS DR8 and 84.45% on SDSS DR17, increasing by 13.65% and 4.04% respectively compared to the single modal case. In addition, the RMS of $|\Delta z|$ is 0.0128 and 0.0097 lower than that in the single modal. These results are comparable to the prediction results obtained by using the two data sets of SDSS-WISE in Zhang et al. (2013), which proves that our method is superior.

In terms of its generality, our model would only require minor changes to perform other tasks, such as galaxy classification and galaxy redshift prediction. For example, the MSE of the prediction result with the method in Section 3.1 can reach 0.0011, which is equivalent to, or even more accurate, than the results of 0.0014 and 0.0019 in Mu et al. (2020). Therefore, our method also has certain generality.

In conclusion, our PhotoRedshift-MML can achieve the best RMS of $|\Delta z|$ and significantly outperform the previously published state-of-the-art work in a single modal since it can enrich the characteristic features by exploiting another simulated modal data. In future research works, we plan to fuse the data of other sky survey plans, such as UKIDSS and WISE, which is expected to further improve the accuracy of multimodal prediction of redshift with the help of more observation data. Furthermore, we believe that the method of multimodal machine learning is a general method, and plan to apply
Table 2. Comparison of quasar photometric redshift prediction results between Baseline 2 and our method.

| Data Set      | Input Pattern | |Δ|z|<0.1(%) | |Δ|z|<0.2(%) | |Δ|z|<0.3(%) | RMS of |Δ|z|
|---------------|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SDSS DR7      | 4C, r         | 78.63±0.23      | 85.70±0.27      | 87.09±0.23      | 0.259±0.003     |
| SDSS DR7-WISE | 6C, i         | 88.64±0.25      | 96.38±0.15      | 97.63±0.10      | 0.117±0.004     |

Results of Baseline 2

| Data Set      | Input Pattern | |Δ|z|<0.1(%) | |Δ|z|<0.2(%) | |Δ|z|<0.3(%) | RMS of |Δ|z|
|---------------|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SDSS DR17     | \( \chi_{\text{phot}} \) | 80.41            | 92.60            | 96.06            | 0.1332          |
| SDSS DR17     | \( \langle \chi_{\text{phot}, \text{spec}} \rangle \) | 84.45            | 93.59            | 96.43            | 0.1235          |
| SDSS DR8      | \( \chi_{\text{phot}} \) | 77.13            | 94.47            | 97.13            | 0.1174          |
| SDSS DR8      | \( \langle \chi_{\text{phot}, \text{spec}} \rangle \) | 90.78            | 95.62            | 97.35            | 0.1046          |

Our results

Table 3. Comparison between single modal and multi-modal method on SDSS DR17.

| Input Pattern | Outlier Fraction | \( \sigma_{NMAD} \) | Bias       |
|---------------|------------------|---------------------|------------|
| \( \chi_{\text{phot}} \) | 0.1050           | 0.0732              | -0.0192    |
| \( \langle \chi_{\text{phot}, \text{spec}} \rangle \) | 0.0905           | 0.0520              | -0.0042    |

it to other tasks, e.g. photometric redshifts estimation of galaxies and classification of celestial objects.

ACKNOWLEDGEMENTS

This work is supported by the science research grants from the National Science Foundation of China (No. U1931209) and the China Manned Space Project with NO.CMS-CSST-2021-B05. The authors are also highly grateful for the constructive suggestions given by Jiali Deng.

DATA AVAILABILITY

The data underlying the research results is available in SDSS DR17, on https://skyserver.sdss.org/CasJobs/ and https://dr17.sdss.org/optical/spectrum/search for photometric data and spectroscopic data, respectively.

REFERENCES

Ansari Z., Aghello A., Gall C., 2021, Astronomy & Astrophysics, 650, A90
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Luechin F., Fontana A., Giallongo E., 1999, Monthly Notices of the Royal Astronomical Society, 310, 540
Ball N. M., Brunner R. J., Myers A. D., Strand N. E., Alberts S. L., Tcheng D., 2008, The Astrophysical Journal, 683, 12
Balrušaitis T., Ahuja C., Morency L.-P., 2018, IEEE transactions on pattern analysis and machine intelligence, 41, 423
Baron D., 2019, arXiv preprint arXiv:1904.07248
Beck R., Dobos L., Budavári T., Szalay A. S., Csabai I., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1371
Bonnett C., 2015, Monthly Notices of the Royal Astronomical Society, 449, 1043
Brammer G. B., van Dokkum P. G., Coppi P., 2008, The Astrophysical Journal, 686, 1503
Brescia M., Cavauto S., D’Abrusco R., Longo G., Mercurio A., 2013, The Astrophysical Journal, 772, 140
Brescia M., Salvato M., Cavauto S., Ananna T., Riccio G., LaMassa S., Urry C., Longo G., 2019, Monthly Notices of the Royal Astronomical Society, 489, 663
Brescia M., Cavauto S., Razim O., Amaro V., Riccio G., Longo G., 2021, arXiv preprint arXiv:2104.06774
Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, The Astrophysical Journal, 712, 511
Carrasco Kind M., Brunner R. J., 2014, Monthly Notices of the Royal Astronomical Society, 442, 3380
Cavauto S., Brescia M., Longo G., Mercurio A., 2012, Astronomy & Astrophysics, 546, A13
Cavauto S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017, Monthly Notices of the Royal Astronomical Society, 465, 1959
Cunha P., Humphrey A., 2022, arXiv preprint arXiv:2204.02080
Curran S., 2020, Monthly Notices of the Royal Astronomical Society: Letters, 493, L70
Dahlen T., et al., 2013, The Astrophysical Journal, 775, 93
Desprez G., et al., 2020, Astronomy & Astrophysics, 644, A31
Duncan K. J., Jarvis M. J., Brown M. J., Röttgering H. J., 2018, Monthly Notices of the Royal Astronomical Society, 477, 5177
D’Isanto A., Polsterer K. L., 2018, Astronomy & Astrophysics, 609, A111
Fluke C. J., Jacobs C., 2020, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10, e1349
Fu Y., Wu X.-B., Yang Q., Brown A. G., Feng X., Ma Q., Li S., 2021, The Astrophysical Journal Supplement Series, 254, 6
Hong S., Zou Z., Xu L., 2021, Acta Astronomica Sinica, 62, 13
Hoyle B., 2016, Astronomy and Computing, 16, 34
Hoyle B., Rau M. M., Bonnett C., Seitz S., Weller J., 2015, Monthly Notices of the Royal Astronomical Society, 450, 305
Ilbert O., et al., 2006, Astronomy & Astrophysics, 457, 841
Ilbert O., et al., 2008, The Astrophysical Journal, 690, 1236
Laurino O., D’Abrusco R., Longo G., Riccio G., 2011, Monthly Notices of the Royal Astronomical Society, 418, 2165
Leistedt B., Hogg D. W., 2017, The Astrophysical Journal, 838, 5
Li X., Liu Z., Hu Z., Wu F., Zhao Y., 2007, Spectroscopy and Spectral Analysis, 27, 4
Li C., et al., 2022, Monthly Notices of the Royal Astronomical Society, 509, 2289
Lyke B. W., et al., 2020, The Astrophysical Journal Supplement Series, 250, 8
Mountrichas G., Corral A., Masoura V., Georgantopoulos I., Ruiz A., Georgakakis A., Carrera F. J., Fotopoulou S., 2017, Astronomy & Astrophysics, 608, A39
Mroueh Y., Marcheret E., Goel V., 2015, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2130–2134
Mu Y.-H., Qiu B., Zhang J.-N., Ma J.-C., Fan X.-D., 2020, Research in Astronomy and Astrophysics, 20, 089
Owens A., Isola P., McDermott J., Torralba A., Adelson E. H., Freeman W. T., 2016, in Proceedings of the IEEE conference on computer vision and pattern recognition. pp 2405–2413

Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, Astronomy & Astrophysics, 621, A26

Peng N., Zhang Y., Zhao Y., 2010, in Software and Cyberinfrastructure for Astronomy. p. 77402T

Peterson B. M., 1997, An introduction to active galactic nuclei. Cambridge University Press

Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, Monthly Notices of the Royal Astronomical Society, 447, 3500

Razim O., Cavuoti S., Brescia M., Riccio G., Salvato M., Longo G., 2021, Monthly Notices of the Royal Astronomical Society, 507, 5034

Sadeh I., Abdalla F. B., Lahav O., 2016, Publications of the Astronomical Society of the Pacific, 128, 104502

Salvato M., Ilbert O., Hoyle B., 2019, Nature Astronomy, 3, 212

Schmidt S., et al., 2020, Monthly Notices of the Royal Astronomical Society, 499, 1587

Scottiz V., et al., 2016, Monthly Notices of the Royal Astronomical Society, 462, 1683

Speagle J. S., Eisenstein D. J., 2017, Monthly Notices of the Royal Astronomical Society, 469, 1186

Wu X.-B., Jia Z., 2010, Monthly Notices of the Royal Astronomical Society, 406, 1583

Wu Z., Jiang Y.-G., Wang J., Pu J., Xue X., 2014, in Proceedings of the 22nd ACM international conference on Multimedia. pp 167–176

Yèche C., et al., 2010, Astronomy & Astrophysics, 523, A14

York D. G., et al., 2000, AJ, 120, 1579

Zhang Y., Ma H., Peng N., Zhao Y., Wu X.-b., 2013, The Astronomical Journal, 146, 22

Zou Z., Zhu T., Xu L., Luo A.-L., 2020, Publications of the Astronomical Society of the Pacific, 132, 044503

**APPENDIX A: SQL QUERY**

```sql
SELECT a.plate, a.mjd, a.fiberID, a.specObjID, a.bestObjID, a.class, a.z, a.zErr, b.objID, b.psfMag_u, b.psfMag_g, b.psfMag_r, b.psfMag_i, b.psfMag_z, b.extinction_u, b.extinction_g, b.extinction_r, b.extinction_i, b.extinction_z, b.psfMagErr_u, b.psfMagErr_g, b.psfMagErr_r, b.psfMagErr_i, b.psfMagErr_z
INTO mydb.dr17psf_ugriz_SNR
FROM dr17.SpecObjAll as a
JOIN dr17.PhotoObjAll as b
ON a.bestObjID = b.objID
WHERE a.class = 'QSO' AND a.z >= 1 AND a.z <= 5 AND a.zErr < 0.001 AND a.snMedian_u < 10 AND a.snMedian_g < 10 AND a.snMedian_r < 10 AND a.snMedian_i < 10 AND a.snMedian_z < 10 AND b.psfMag_g >= 18 AND b.psfMag_g <= 22 AND b.petroRad_r < 5
```

This paper has been typeset from a \TeX/\LaTeX\ file prepared by the author.