The Early-type Stars from the LAMOST Survey: Atmospheric Parameters

Yanjun Guo (郭彦君)1,2,3, Bo Zhang4, Chao Liu2,5, Jiao Li5, Jianguan Li5,2, Luqian Wang1,2, Zhicun Liu2, and Xuefei Chen1,2

1 Yunnan Observatories, Chinese Academy of Sciences, Kunming, 650216, People’s Republic of China; 2 School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing, 100049, People’s Republic of China; 3 Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China; 4 Department of Astronomy, Beijing Normal University, Beijing 100875, People’s Republic of China; 5 Nanjing Institute of Astronomical Optics & Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, People’s Republic of China

Abstract

Massive stars play key roles in many astrophysical processes. Deriving the atmospheric parameters of massive stars is important for tracing the evolution of massive stars. Here we report our work on adopting the data-driven technique called stellar label machine (SLAM) with the nonlocal thermal equilibrium TLUSTY synthetic spectra as the training data set to estimate the stellar parameters of Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) optical spectra for early-type stars. We apply two consistency tests to verify this machine-learning method and compare stellar labels given by SLAM with the labels in the literature for several objects having high-resolution spectra. We provide the stellar labels of effective temperature \( T_{\text{eff}} \), surface gravity \( \log g \), metallicity \([M/H]\), and projected rotational velocity \((v \sin i)\) for 3931 and 578 early-type stars from the LAMOST low-resolution survey (LRS) and medium-resolution survey (MRS), respectively. To estimate the average statistical uncertainties of our results, we calculated the standard deviation between the predicted stellar label and the prelabeled published values from the high-resolution spectra. The uncertainties of the four parameters are \( \sigma(T_{\text{eff}}) = 2185 \text{ K}, \sigma(\log g) = 0.29 \text{ dex}, \) and \( \sigma(v \sin i) = 11 \text{ km s}^{-1} \) for MRS, and \( \sigma(T_{\text{eff}}) = 1642 \text{ K}, \sigma(\log g) = 0.25 \text{ dex}, \) and \( \sigma(v \sin i) = 42 \text{ km s}^{-1} \) for LRS spectra, respectively. We note that the parameters of \( T_{\text{eff}}, \log g, \) and \([M/H]\) can be better constrained using LRS spectra than using MRS spectra, most likely due to their broad wavelength coverage, while \( v \sin i \) is constrained better by MRS spectra than by LRS spectra, probably due to the relatively accurate line profiles of MRS spectra.

Unified Astronomy Thesaurus concepts: Early-type stars (430); Astronomy data analysis (1858); Surveys (1671); Catalogs (205)

Supporting material: machine-readable tables

1. Introduction

Early-type stars are massive and luminous, and they are mainly comprised of O- or B-type stars (Morton & Adams 1968; Morgan & Keenan 1973; Panagia 1973). Massive early-type stars are possible progenitors of extremely compact stellar objects, such as neutron stars, black holes, high-mass X-ray binaries, and Type Ib/c supernovae, and they are potential sources of gravitational wave events (Sadowski et al. 2008; Yoon et al. 2010; Abbott et al. 2016a, 2016b; Chen et al. 2018; Han et al. 2020; Langer et al. 2020). Steller atmospheric parameters, such as effective temperature \( T_{\text{eff}} \), surface gravity \( \log g \), projected rotational velocity \((v \sin i)\), and metallicity \([M/H]\), are often referred to as the stellar labels. Deriving such labels for massive stars is important to reveal their physical properties and to constrain their location in the Hertzsprung–Russell (HR) diagram, and this information is an essential component for tracing the evolutionary scenario of massive stars (Fitzpatrick & Garmany 1990; Langer & Maeder 1995; McErlean et al. 1999). The spectral chemical analysis of massive stars could be used to determine the present-day abundance of local and external galaxies (Daflon et al. 2001; Urbanaje et al. 2005; Esteban et al. 2017). Early-type stars are also important for studies of cosmic ionization and act as the major sources of energetic feedback for the interstellar medium and intergalactic medium (Hopkins et al. 2014).

The classical way to derive stellar labels of massive stars is by comparing observations to theoretical stellar atmospheric models through \( \chi^2 \) minimization approaches. We summarize several works on determining the stellar labels of early-type stars using such techniques as follows. Trundle et al. (2007) investigated stellar labels of 61 B-type stars in the Galaxy and Magellanic Clouds using high-resolution spectra from the Very Large Telescope (VLT). Hunter et al. (2009) obtained the surface nitrogen abundance and projected rotational velocity for about 150 B-type stars located in the Galaxy and Small Magellanic Clouds using the spectrophotometric observations from the VLT-FLAMES survey. A comprehensive study of the early B-type stars located within the solar neighborhood was carried out by Nieva & Przybilla (2012) using multiple sets of high-resolution spectra. McEvoy et al. (2017) measured atmospheric parameters and metallic abundance for a sample of runaway B-type stars in the Galaxy to trace their formation mechanism. These studies above all used high signal-to-noise ratio (S/N) spectra that cover wavelength regions that are rich in atmospheric absorption lines to proceed with the...
line diagnostic analysis, but the investigations were limited to a small sample of observations.

The alternative way of measuring stellar labels for stellar systems is a data-driven approach. This technique learns the prelabeled reference spectra and forms a model that maps the stellar labels to spectra. The data-driven technique is widely used and has demonstrated reliable applicability in many astronomical studies. For example, Ness et al. (2015) constructed the “The Cannon” model using a large set of observed spectra for reference stars with prior known stellar labels, and this model was applied to spectra observed from the APOGEE DR10 to estimate the stellar labels for about 55,000 stars. The predicted results are in good agreement with those obtained from the APOGEE pipeline. Xiang et al. (2020) employed a neural network technique using the low-resolution spectra from Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) DR5 to estimate the absolute magnitude for a sample of 16,002 OB stars. Zhang et al. (2020a, 2020b) employed a neural network technique using the low-resolution spectra from the APOGEE DR15 spectral library to derive stellar labels for K-giant stars. More recently, Li et al. (2021) adopted the SLAM to derive stellar labels for M dwarfs using spectral observations from the LAMOST medium-resolution survey (LAMOST-MRS) database.

Only a few early-type star surveys are available (Barbá et al. 2010; Maíz Apellániz 2010; Simón-Díaz et al. 2011). However, the stellar labels derived using the spectra from these surveys vary from case to case, mainly because the spectral observations were made with different instruments. Consequently, different pipelines and methods were applied to analyze the spectral properties of the data (Sana 2017). Catalogs of stellar labels predicted from a large sample of homogeneous spectral observations using a consistent method are critically lacking. Motivated by the recent release of a large number of spectral observations from the LAMOST database, in this paper, we report our work of constructing a catalog consisting of homogenized stellar atmospheric parameters for early-type stars identified from the survey by applying the data-driven learning module SLAM. This catalog is the first that consists of consistently derived stellar labels for such a large sample of early-type stars, and this information will be helpful to provide a reference for studying the physical and evolutionary properties of massive stars.

In the following text, we describe the sample selection of the LAMOST spectra in Section 2. We briefly describe the algorithm of the data-driven learning module SLAM, the consistency check of applying the module to the observational data, and details of constructing the training data set in Section 3. We report our results of predicting stellar labels for the LAMOST spectra in Section 4. We discuss our results in Section 5 and summarize the conclusion in 6.

2. LAMOST Data Selection

LAMOST is a 4 m quasi-meridian reflecting Schmidt telescope located at the Xinlong station of the National Astronomical Observatory. The telescope has a field of view of 20 square degrees and is implemented with 4000 fibers. Both a low-resolution spectrograph and a medium-resolution spectrograph were installed on the telescope. The low-resolution spectrograph has a resolving power of $R \sim 1800$ with a wavelength coverage within a range of 3690–9100 Å; about ∼9 million optical spectra are released at the time of writing (Cui et al. 2012; Deng et al. 2012; Zhao et al. 2012). The medium-resolution spectrograph has $R \sim 7500$ and includes both a blue arm (B) and a red arm (R). Spectral observations made from the blue arm span a wavelength in the range of 4950–5350 Å, and the spectra observed using the red arm cover the wavelength range of 6300–6800 Å (Liu et al. 2020). LAMOST finished the first observing campaign from 2011 to 2018 (known as LAMOST I). The second active survey program (LAMOST II) began in 2017 September. About five million medium-resolution spectra observed with a single exposure for ∼800,000 stars were collected by 2019 June (Liu et al. 2020).

Liu et al. (2019) identified 16,032 early-type stars through measuring the equivalent widths of several absorption line profiles using low-resolution spectra from the LAMOST low-resolution survey (LRS) database. Guo et al. (2021) used the spectra from the LAMOST MRS database to investigate the multiplicity properties of early-type stars (hereinafter Paper I). In that work, a total number of 9382 early-type stars were identified by measuring the equivalent widths of Hα λ6564 Å, He I λ6678 Å, and Mg I profiles in the range of 5167–5184 Å. We display the distribution of the two samples in Figure 1, in which the gray dots represent early-type stars from the LAMOST-LRS, and red dots form LAMOST-MRS. In this study, we adopt all of the identified O- and B-type stars from Liu et al. (2019) and Paper I to derive the stellar labels of their spectra by applying the SLAM.

3. Method

3.1. SLAM

The SLAM (Zhang et al. 2020a, 2020b) is a forward stellar spectral model based on the support vector regression (SVR) algorithm (Cortes & Vapnik 1995), which is widely applied in spectral analysis. A comprehensive discussion of the algorithm and architecture of the SLAM are described in Zhang et al. (2020a). The SLAM code is written in python and is available for public downloading from the GitHub website. In order to apply the SLAM to observed spectral data, we adopted several default hyperparameters as suggested by Zhang et al. (2020a) to describe the model complexities in the SVR algorithm. These include the penalty level $(C)$ within a range of $[0.1, 1, 10]$, the width of the radial basis function $(\gamma)$ under the range of $[0.1, 1]$, and a tube radius $(\rho)$ with a value of 0.05.

In principle, three steps are included to apply the SLAM to observational spectra data, and we briefly summarize these as follows:

Step 1: Preprocessing. We first need to standardize the spectra in the training set such that their stellar labels and spectral fluxes have a mean value of 0 and a variance with a value of 1. This step aims to map the stellar labels of spectra in the training set and normalize the spectra into standardized space. This procedure is automatically achieved by SLAM (Zhang et al. 2020a) through an internal computational algorithm.

Step 2: Training. We use the spectra in the training set to train the SVR model such that it learns the knowledge of the prelabeled spectra from the input spectral data themselves, and then it maps the corresponding atmospheric parameters and spectra to a model.

Step 3: Prediction. Based upon the training procedure as described above, we predict the stellar labels for the observed spectral data using the SVR model.

---

9 http://dr5.lamost.org/

8 https://github.com/hypergravity/astroslam
3.2. The Training Set

In order to apply the SLAM to spectra of early-type stars identified from the LAMOST database, we first need to construct a training set. We search in the literature to collect OB stars with derived stellar labels, but the samples are inadequate to provide complete coverage of the parameter space needed for stellar labels. Therefore, we adopt the TLUSTY, a synthetic spectral library based on the nonlocal thermal equilibrium (NLTE) model as our training set (Lanz & Hubeny 2003, 2007). The NLTE model includes model grids for both O- and B-type stars, including more than 40 elements in the calculations, such as the major atomic species of H, He, C, N, O, Ne, Si, P, S, Fe, and Ni (Lanz & Hubeny 2003, 2007). The solar abundances from Grevesse & Sauval (1998) are adopted in the model. The O-type model spectra assume a microturbulent velocity \( v_\text{t} \) of 10 km s\(^{-1} \), the spectra were sampled irregularly with a sample size of \( 180,000 \sim 200,000 \) pixels, and the wavelength spans from 3000 to 7500 Å. For B-type model spectra, about 175,000 pixel points were used to sample the model grids in irregular format with \( v_\text{t} = 10 \) km s\(^{-1} \). In addition, \( v_\text{t} = 2 \) km s\(^{-1} \) was introduced to the spectra, and the model grids were sampled with 380,000 points. The B-type model grids cover a wavelength range of 3200 ~ 10,000 Å.

We adopted all the available O- and B-type model spectra from the grids. The O-type model grids include \( T_{\text{eff}} \) spanning a range of 27,500 ~ 55,000 K with a step size of 2500 K, and \( \log g \) within a range of 3.00 ~ 4.75 dex with a step size of 0.25 dex. We chose \( T_{\text{eff}} \) within a range of 15,000 ~ 30,000 K with a step size of 1000 K, and \( \log g \) within a range of 1.75 ~ 4.75 dex with a step size of 0.25 dex for B-type stars. The metallicity \( [M/H] \) with values between -1.0 and 0.3 dex was chosen for both O-type and B-type model spectra. Projected rotational velocity and macroturbulent velocities were not included in all of these spectra. In Figure 2 we show the TLUSTY model grids for both O-type stars (crosses) and B-type stars (circles) in the \( (T_{\text{eff}}, \log g) \) plane.

We downloaded the synthetic model O-type spectra\(^9\) and the B-type spectra\(^10\) from the TLUSTY websites. In order to increase the sample size, we randomly generated the model spectra using stellar parameters within the available range given by the TLUSTY grids through a linear interpolation approach, and these expanded grids are shown as the gray shaded area in Figure 2. We then randomly convolved these model spectra with projected rotational velocity \( v \sin i \) with values within the range of 0 ~ 300 km s\(^{-1} \). In order to bring the generated model spectra into an agreement with those of observations obtained from the LAMOST, we then degraded the resolutions of TLUSTY model spectra down to \( R \sim 1800 \) (comparable to LAMOST-LRS, defined as TLUSTY-LRS) and \( R \sim 7500 \) (comparable to LAMOST-MRS, defined as TLUSTY-MRS) through a Gaussian smoothing using the python package laspec.qconv.\(^11\) We then resampled the wavelength grids of both TLUSTY-LRS and TLUSTY-MRS spectra such that they have a comparable coverage as the observations. The wavelength spans the range of 3900 ~ 7000 Å with a step size of 1 Å for TLUSTY-LRS spectra, and 4950 ~ 5350 Å and 6300 ~ 6800 Å for the red and blue arms of the TLUSTY-MRS spectra with a step size of 0.2 Å, respectively. We excluded bad pixels from the spectra, and the training set consists of 1000 TLUSTY-LRS spectra and 5000 TLUSTY-MRS spectra (see Section 3.4).

\(^9\) http://tlusty.oca.eu/Tlusty2002/tlusty-frames-OS02.html
\(^10\) http://tlusty.oca.eu/Tlusty2002/tlusty-frames-BS06.html
\(^11\) https://github.com/hypergravity/laspec/
that stellar labels of spectra in the validation set will be

declared in Section 3.3, we added the set of \( N \) values to each pixel of the TLUSTY-LRS/MLRS spectra using a Gaussian function to evaluate the performances of

\[ \text{SLAM} \] in different S/Ns. The S/N values span a range of

\( 20 \sim 100 \) with an increment of 20.

In principle, a decreasing trend is shown from the CV-bias and CV-scatter tests, indicating that the data-driven method is suitable. In Figure 3 we show the distributions for both the CV-scatter and the CV-bias values for each predicted stellar label as a function of S/N. In panels (a) to (d), the CV-scatter values calculated for \( T_{\text{eff}}, \log g, [\text{M/H}], \text{and} \ v \sin \ i \) parameters using the TLUSTY-LRS (red lines with squares) and TLUSTY-MRS (blue lines with stars) spectra all display a decreasing trend toward a higher S/N. The calculated CV-bias values in panels (a)–(d) all approach zero toward higher S/N values, indicating that the stellar labels predicted by the SLAM for spectra in the validation set are consistently matched with the prelabeled values as given from the TLUSTY model grids.

3.4. Determining the Sample Sizes and SLAM Errors

In principle, enlarging the size of the training sample by including more spectra results in smaller uncertainties of the stellar labels predicted from the SLAM, but the computation costs a significant amount of time. We thus need to determine the size of the training sample such that the precision of the prediction task can be combined with a reasonable computing time budget. In addition to the original training set consisting of 1000 TLUSTY-LRS model spectra, we constructed an extra training set including 5000 model spectra. In addition to the exiting training set of 5000 TLUSTY-MRS model spectra, we generated two additional sets consisting of 1000 and 10,000 model spectra. By following the procedures as described in Section 3.3, we added the set of S/N values to each pixel of the model spectra in the individual training set for both TLUSTY-LRS and TLUSTY-MRS and then predicted their associated stellar labels. The SLAM errors for each of the predicted stellar labels are obtained by computing the standard deviation between the predicted values from the SLAM and the true labels as given from the TLUSTY model grids.

Before we apply the SLAM to the identified early-type stars from the LAMOST database to predict their stellar labels, we first need to verify the robustness of applying the module to the training set. The verification is achieved through the usage of a consistency check, called cross-validation (CV). CV is also called \( k \)-fold cross-validation. It is a popular method that is used to score the performance of machine-learning models in a less biased or less optimistic approach. This approach divides the input spectra into \( k \) groups. Among them, the \( k-I \) groups are randomly selected as the training set, and the spectra that remained in the \( k \) set are chosen to be the validation set, such that stellar labels of spectra in the validation set will be predicted using the SLAM.

We adopted the five-fold CV technique from Ojala & Garriga (2009). We use CV-scatter (standard deviation) and CV-bias (mean deviation) to describe the performance of CV, and the equations are given as follows:

\[
\text{CV-scatter} = \frac{1}{n} \sum_{i=1}^{n} (\theta_{i, \text{SLAM}} - \theta_i)^2, \tag{1}
\]

and

\[
\text{CV-bias} = \frac{1}{n} \sum_{i=1}^{n} (\theta_{i, \text{SLAM}} - \theta_i), \tag{2}
\]

where \( \theta_i \) is the true stellar label of the star with an index of \( i \).

\( \theta_{i, \text{SLAM}} \) denotes the stellar label predicted from the SLAM of the same star.

The values of CV-scatter and CV-bias are dependent on the signal-to-noise ratio, S/N, of the input spectra. We thus investigate the distribution of the CV values as a function of the S/N added to the spectra. We added a set of S/N values to each pixel of the TLUSTY-LRS/MLRS spectra using a Gaussian function to evaluate the performances of

\[ \text{SLAM} \] in different S/Ns. The S/N values span a range of

\( 20 \sim 100 \) with an increment of 20.

In principle, a decreasing trend is shown from the CV-bias and CV-scatter tests, indicating that the data-driven method is suitable. In Figure 3 we show the distributions for both the CV-scatter and the CV-bias values for each predicted stellar label as a function of S/N. In panels (a) to (d), the CV-scatter values calculated for \( T_{\text{eff}}, \log g, [\text{M/H}], \text{and} \ v \sin \ i \) parameters using the TLUSTY-LRS (red lines with squares) and TLUSTY-MRS (blue lines with stars) spectra all display a decreasing trend toward a higher S/N. The calculated CV-bias values in panels (a)–(d) all approach zero toward higher S/N values, indicating that the stellar labels predicted by the SLAM for spectra in the validation set are consistently matched with the prelabeled values as given from the TLUSTY model grids.
shown from the different sample sizes. We thus chose a sample size of 5000 spectra for the TLUSTY-MRS training set.

4. Results and Validation

After learning the spectral features using the model spectra of TLUSTY-LRS and TLUSTY-MRS, now we are ready to apply the SLAM to observational spectra for early-type stars from the LAMOST database to predict their stellar labels. We downloaded a collection of 16,032 LAMOST-LRS spectra and 9382 LAMOST-MRS early-type star spectra from the LAMOST archive sites. The archived spectra were reduced, and wavelength calibration was applied following the pipeline as described in Luo et al. (2015). We then used the python package laspec (see footnote 13) to normalize the spectra using a spline fit. An example of an LAMOST-MRS spectrum (black lines) with OBSID of 703901108 and the associated spline fit (red lines) is shown in the upper panels of Figure 5. The normalization procedure for the same star but with an LAMOST-LRS observation is shown in the upper panel of Figure 6. Before applying the SLAM to the observational data, we first need to bring the LAMOST spectra into the rest frame such that they are in agreement with those of the TLUSTY model spectra. We achieved this goal by collecting the radial velocity (RV) measurements of our sample stars that appear in the work of Zhang et al. (2021), who performed the RV measurements for all stars in the LAMOST DR7 through a cross-correlation algorithm. We then corrected the RV variations in our observed spectra. In order to avoid confusion in analyzing the spectral features using the model spectra of TLUSTY-LRS and TLUSTY-MRS, we use a smoothing spline to fit the observed spectrum and clip the pixels $3\sigma$ above or $1\sigma$ below the median of the residual and iterate this process three times to avoid strong lines, which is equivalent to a low-pass filtering. We obtain the pseudocountinuum from the remaining smoothed pixels, and then the obtained normalized spectrum by dividing the observed spectrum by the pseudocountinuum.

Figure 3. The distributions of the CV-scatter (solid lines) and the CV-bias (dashed lines) values for predicted stellar labels of $T_{\text{eff}}$ (panel (a)), $\log g$ (panel (b)), $[M/H]$ (panel (c)), and $v\sin i$ (panel (d)) as a function of S/N values. Results obtained from the TLUSTY-LRS spectra and the TLUSTY-MRS spectra in the training set are shown as squares and stars, respectively.

12 http://dr5.lamost.org/
13 http://dr7.lamost.org/
14 We use a smoothing spline to fit the observed spectrum and clip the pixels $3\sigma$ above or $1\sigma$ below the median of the residual and iterate this process three times to avoid strong lines, which is equivalent to a low-pass filtering. We obtain the pseudocountinuum from the remaining smoothed pixels, and then the obtained normalized spectrum by dividing the observed spectrum by the pseudocountinuum.
The LAMOST-LRS spectrum of the same star has predicted stellar labels of $T_{\text{eff}} = 16,494$ K, $\log g = 4.2$ dex, $[M/H] = -0.30$ dex, and $v \sin i = 4$ km s$^{-1}$ as given by SLAM, and the model spectrum with the estimated stellar labels is shown in Figure 6.

### 4.1. The Stellar Parameters of OB Stars

By applying the SLAM to the sample of stars selected from the LAMOST database, we predict their stellar labels. In Table 1 we report the spectral observational ID, the equatorial coordinates, $S/N$ values, the predicted values of $T_{\text{eff}}$, the $\log g$, the $[M/H]$, and $v \sin i$ of 16,032 early-type stars from LAMOST-LRS. In the sample, the stellar labels of 3931 stars are within the parameter ranges given by the training set. We assigned a flag of “I” to these stars, and they are listed in Column 9 of Table 1. For the remaining stars outside the parameter range, we obtained their stellar labels by extrapolating values from the SLAM. We retained the predicted stellar labels of these stars in Table 1 as preliminary estimates for the stellar parameters. In Table 2 we list the observational ID, the equatorial coordinates, $S/N$ values, the predicted stellar labels of $T_{\text{eff}}$, the $\log g$, the $[M/H]$, and $v \sin i$, the observational date (MJD), $S/N$ R (red arm), $S/N$ B (blue arm), and the index of flags for 9238 early-type stars from the LAMOST-MRS database. In the sample, 578 stars show stellar labels within the parameter ranges of the associated training set.

### 4.2. Hertzsprung–Russell Diagram

In Figure 7 we show the density distributions of predicted $T_{\text{eff}}$ and $\log g$ values for both LAMOST-LRS and LAMOST-MRS spectra. The vertical color bar on the right represents the number density of the stars, and the blue box represents the restricted parameter ranges of $T_{\text{eff}}$ and $\log g$ selected from the TLUSTY model grids (see Section 3.2). We derived the isochrones from the Padova and Trieste Stellar Evolutionary Code (PARSEC) (Hurley et al. 2000; Bressan et al. 2012), and overplotted them with the ages of 1, 10, and 100 Myr in the figure. Two sets of isochrones with values of $[M/H] = -1.0$ dex (green lines) and $[M/H] = 0.0$ dex (blue lines) are shown for each of the age tracks. The theoretical isochrones are well fit with the predicted stellar labels for the observed spectra as given by the SLAM.

We see in Figure 7 that a large fraction of the stars in the sample with predicted stellar labels are located outside the restricted parameter range of our training set. Based upon their location on the plot, they can be divided into three groups, and we briefly discuss their features below. The first group of stars shows an estimated $T_{\text{eff}} < 15,000$ K (with spectral type later than B5). This suggests that late B-type stars dominate the
population of our sample of LAMOST early-type stars, and this finding is consistent with the result suggested by the initial mass function, i.e., the population of late-type stars shows an increasing trend toward low stellar mass. In the upper panel of Figure 7, a second group of stars consists of 363 LAMOST-LRS spectra showing predicted values of $g \log 2.0 cm^{-2}$. We visually inspected the spectra of these stars and found that they are either featureless or are dominated by the presence of the Hα emission profile. We caution that these features that appear in the spectra might lead to an underestimation of the true $g \log$ values of the observations. The last group of stars is found with predicted values of $g \log 4.75 cm^{-2}$, and they are likely subdwarf-type stars, which have typical $g \log$ values in the range of 5.1–6.4 cm s$^{-2}$ (Heber 2009).

### 4.3. Comparison to High-resolution Spectra

We also predict the stellar labels for a sample of prelabeled high-resolution spectra (HRS) from publications to verify the consistency of our training set. We select a sample of 28 early-type stars with preestimated stellar labels from works of Trundle et al. (2007), Nieva & Przybilla (2012), and McEvoy et al. (2017). In order to directly compare the results of these archived HRS to those of the LAMOST-LRS and LAMOST-MRS data, we first degraded the resolution of the HRS down to resolutions of $R \sim 1800$ (HRS-LRS) and $R \sim 7500$ (HRS-MRS). We then applied the SLAM to both of the degraded HRS-LRS and HRS-MRS spectra to predict their stellar label values, including $T_{\text{eff}}$, $\log g$, and $v \sin i$. A comparison of our predicted stellar labels using the SLAM to the prelabeled values obtained from publications is shown in Figure 8. The left and middle panels indicate that the predicted $T_{\text{eff}}$ and $\log g$ values from the HRS-LRS spectra (red circles) agree with the published estimates of HRS within the uncertainty range. These two parameters are better constrained than those of the HRS-MRS spectra (blue squares). The right panel shows that the predicted $v \sin i$ values of the HRS-MRS spectra are
Table 1
The Predicted Stellar Labels of 16,052 LAMOST-LRS Early-type Stars

| Star OBSID | R.A. (deg) | Decl. (deg) | S/N | $T_{\text{eff}}$ (K) | log g (dex) | $[\text{M}/\text{H}]$ (dex) | $v \sin i$ (km s$^{-1}$) | Flag |
|-----------|-----------|------------|-----|-----------------|------------|----------------|----------------|-----|
| 51401075  | 237.0048  | −2.5811    | 41  | 39,714          | 4.4        | −0.74          | 6              | I   |
| 203808205 | 93.7723   | 12.3566    | 39  | 31,784          | 4.2        | −0.60          | 8              | I   |
| 253216223 | 322.9148  | 50.8778    | 21  | 37,341          | 4.0        | −0.56          | 1              | I   |
| 255008029 | 81.8724   | 36.4042    | 45  | 18,063          | 4.1        | −0.48          | 3              | I   |
| 360106063 | 297.1958  | 44.3561    | 31  | 24,675          | 4.3        | −0.09          | 13             | I   |
| 536607134 | 95.1156   | 21.8162    | 51  | 15,550          | 4.3        | −0.13          | 5              | I   |

(This table is available in its entirety in machine-readable form.)

Table 2
The Predicted Stellar Parameters of 9238 LAMOST-MRS Early-type Stars

| Star OBSID | R.A. (deg) | Decl. (deg) | Date (MJD) | S/N$_N$ | S/N$_R$ | $T_{\text{eff}}$ (K) | log g (dex) | $[\text{M}/\text{H}]$ (dex) | $v \sin i$ (km s$^{-1}$) | Flag |
|-----------|-----------|------------|------------|---------|---------|-----------------|------------|----------------|----------------|-----|
| 684408119 | 35.4066   | 57.2101    | 58,421.0472 | 97      | 80      | 25,644          | 4.6        | −0.25          | 112            | I   |
| 627415119 | 36.1931   | 57.4642    | 58,119.8188 | 272     | 190     | 22,287          | 2.8        | −0.75          | 131            | I   |
| 692513076 | 48.8320   | 65.9180    | 58,444.9562 | 101     | 69      | 17,908          | 4.4        | −0.25          | 75             | I   |
| 729206077 | 84.0657   | 34.2383    | 58,539.8562 | 97      | 56      | 26,271          | 4.4        | −0.09          | 36             | I   |
| 609116071 | 18.4606   | 59.7916    | 58,088.8562 | 95      | 75      | 22,445          | 3.8        | −0.10          | 5              | I   |
| 591515050 | 19.3419   | 58.8882    | 58,030.9861 | 159     | 131     | 24,700          | 3.7        | −0.10          | 44             | I   |

(This table is available in its entirety in machine-readable form.)

Figure 7. The distribution of predicted stellar labels of early-type stars for both LAMOST-LRS (upper panel) and LAMOST-MRS (bottom panel) spectra is shown in the ($T_{\text{eff}}, \log g$) plane. The vertical color bars in both panels represent the number density of the stars. The blue box represents the restricted parameter range of $T_{\text{eff}}$ and log g values selected from the TLUSTY model grids. The green and blue lines represent PARSEC, the theoretical isochrone tracks, with metallicity values of $[\text{M}/\text{H}] = -1.0$ dex and $[\text{M}/\text{H}] = 0.0$ dex, respectively. A set of tracks with ages of 1, 10, and 100 Myr is included in the figure.

likely were flattened during the spectral degradation process, resulting in a lower estimate for the $v \sin i$.

In Section 3.4 we discuss the errors of the predicted stellar labels given by SLAM. Assuming an $S/N = 100$ is given for input TLUSTY model spectra, based on the distribution of SLAM errors of stellar labels as a function of $S/N$ values shown in Figure 4, we would expect to obtain estimated errors of $\sigma(T_{\text{eff}}) = 327$ K, $\sigma(\log g) = 0.03$ dex, and $\sigma(v \sin i) = 4$ km s$^{-1}$ for TLUSTY-LRS spectra. For a training set consisting of model spectra of TLUSTY-MRS, we obtained errors of $\sigma(T_{\text{eff}}) = 562$ K, $\sigma(\log g) = 0.04$ dex, and $\sigma(v \sin i) = 1$ km s$^{-1}$. We caution that these estimates from the model spectra may underestimate the true values of observational data. We thus calculated the standard deviation as realistic errors between the predicted stellar label and the prelabeled published values from the HRS for each stellar parameter as the realistic errors of the labels. We then arrived at realistic error estimates of $\sigma(T_{\text{eff}}) = 2185$ K, $\sigma(\log g) = 0.29$ dex, and $\sigma(v \sin i) = 11$ km s$^{-1}$ for HRS-MRS spectra, and $\sigma(T_{\text{eff}}) = 1642$ K, $\sigma(\log g) = 0.25$ dex, and $\sigma(v \sin i) = 42$ km s$^{-1}$ for HRS-LRS spectra. These error estimates likely reflect the true uncertainties of the stellar labels derived from the SLAM.

4.4. Comparing the Performance of LRS and MRS from SLAM

In order to assess the dependence of the predicted stellar labels on the spectral resolution using the SLAM, we cross-matched the early-type stars listed in Liu et al. (2019) (LAMOST-LRS) with those of Paper I (LAMOST-MRS). A sample of 229 stars in common with an $S/N$ value greater than 100 was found. In Figure 9 we compare the predicted stellar labels from the LAMOST-MRS spectra ($y$-axis) to those from the LAMOST-LRS ($x$-axis). The gray region in each panel of Figure 9 represents the restricted parameter range of the stellar parameters.
selected from the TLUSTY model grids (see Section 3.2). A small sample of stars is located outside the regions (see Section 4.2). We find that most of the predicted stellar labels of $T_{\text{eff}}$, log $g$, and $v \sin i$ values derived from the LAMOST-MRS spectra show higher values than those from the LAMOST-LRS.

Based upon the results shown from the validation tests discussion in Sections 3.3, 3.4, and 4.3, we conclude that the stellar labels of $T_{\text{eff}}$, log $g$, and [M/H] obtained from LAMOST-LRS spectra are better constrained than those of LAMOST-MRS spectra for early-type stars. Similar results are obtained from the work of Zhang et al. (2020b), who employed the SLAM model to predict the stellar labels for F-, G-, and K-type stars using spectra from LAMOST. It is likely that the LAMOST-LRS spectra have a broad
wavelength coverage of 3900–7000 Å compared to that of LAMOST-MRS spectra, and more Balmer and He I line series are included in the spectra, resulting in a better constraint on $T_{\text{eff}}$ and $\log g$. Due to the higher resolution of LAMOST-MRS spectra with an $R \sim 7500$ compared to an $R \sim 1800$ for LAMOST-LRS spectra, more detailed spectral information, such as the outline and shape of broad absorption line profiles, are encoded in the MRS spectra. This results in a better estimation of the projected rotational velocity. We suspect that $v \sin i$ is sensitive to the spectral resolution.

5. Discussion

The stellar labels of massive stars are enclosed in their spectra, and the stellar parameters are often interrelated. For example, for B-type stars, the neutral helium and hydrogen lines are popular lines for estimating the effective temperature, while the hydrogen lines are sensitive to the surface gravity (Dufton et al. 1999). Traditionally, the stellar labels are determined individually through the minimization technique by comparing observed spectra to model spectra, while the SLAM predicts the stellar labels of input spectra simultaneously.

In order to verify the potential degeneracy existing among the predicted stellar labels obtained from the SLAM, we applied a Markov Chain Monte Carlo (MCMC) simulation to inspect the posterior distribution of the stellar labels. We show the results obtained from a LAMOST-MRS spectrum with the observational ID of OBSID 682701103 observed on the night of MJD 58,470.9540 in Figure 10. In the six off-diagonal panels, the contours with 68%, 95%, and 99% confidence regions are plotted as solid black lines from the inside out. In the four diagonal panels, the two vertical dashed lines represent the lower and upper errors of the distribution at the 16th and 84th percentiles, respectively. The distribution indicates that a weak correlation is found between $T_{\text{eff}}$ and $\log g$. However, the variation of the determined stellar labels is sensitive to the spectral resolution.
labels is within the uncertainty range estimated from the SLAM, and we thus consider that the predicted results are still acceptable.

6. Conclusion

Massive stars are important contributors to many astronomical mechanisms. Determining their fundamental physical parameters is vital for understanding the evolutionary scenario of the massive stars. Hitherto, the estimate of stellar parameters for such stars was restricted to small collections of spectral observations. An estimation of the stellar labels using a comprehensive and consistent approach for a large sample of early-type stars is lacking. Motivated by the recent release of a large number of spectra for early-type stars from the LAMOST database, we thus conduct this study to predict stellar labels for more than 20,000 early-type stars identified from the database.

We adopted the synthetic NLTE atmospheric spectral library TLUSTY to generate training sets of TLUSTY-LRS and TLUSTY-MRS model spectra through linear interpolation. The model spectra encompass a wide parameter range to include spectral features of early-type stars over the wavelength regime covered by the LAMOST observations. We then applied a fivefold CV technique to validate the performance of the training sets. By investigating the distribution of SLAM errors of predicted stellar labels as a function of S/N for input training sets with different sample sizes, we determined a sample size of 1000 spectra for the TLUSTY-LRS training set and 5000 spectra for the TLUSTY-MRS training set. We predicted the stellar labels of identified early-type stars from the LAMOST database using a sample of 3931 stars from the LAMOST-LRS and 578 stars from the LAMOST-MRS.

In order to estimate the realistic errors of predicted stellar labels obtained from the SLAM, we collected high-resolution spectra of 28 stars from publications. By downgrading the resolution to match the resolutions of the observed LAMOST-LRS and LAMOST-MRS spectra, we then predicted the stellar labels of these spectra and determined the uncertainties of the associated label by computing the standard deviation between the predicted values from the SLAM and literature values. We then arrived at realistic error estimates of $\sigma(T_{\text{eff}}) = 2185 K$, $\sigma(\log g) = 0.29$ dex, and $\sigma(v\sin i) = 11 \text{ km s}^{-1}$ for HRS-MRS spectra, and $\sigma(T_{\text{eff}}) = 1642 K$, $\sigma(\log g) = 0.25$ dex, and $\sigma(v\sin i) = 42 \text{ km s}^{-1}$ for HRS-LRS spectra. The predicted stellar labels of $T_{\text{eff}}$ and $\log g$ obtained from the MRS spectra are better constrained than those from the LRS spectra, and this is likely due to the broader wavelength coverage of the LRS spectra. However, MRS spectra are more sensitive to constrain the $v\sin i$ values due to the inclusion of detailed line profile features as a result of higher resolutions.

We have demonstrated the applicability of applying the data-driven technique, SLAM, to predict stellar labels for a large set of observations for early-type stars. This technique is promising for a prediction of stellar labels for other types of stars.

This work is supported by the Natural Science Foundation of China (Nos. 11733308, 12090040, 12090043, 11521303, 12125303), by the Yunnan province, by the National Ten-thousand talents program. C.L. acknowledges the National Key R & D Program of China No. 2019YFA0405500 and the NSFC with grant No. 11835057. This work is also supported by the Key Research Program of Frontier Sciences, CAS, grant No. QYZDY-SSW-SLH007 and the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A10.

The Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope 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 the Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

ORCID iDs

Yanjun Guo (郭彦君) (http://orcid.org/0000-0001-9989-9834)
Chao Liu (https://orcid.org/0000-0002-6434-7201)

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, PhRvL, 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, PhRvL, 116, 241103
Barbá, R. H., Gamen, R., Arias, I. J., et al. 2010, RMxAC, 38, 30
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Chen, X., Li, Y., & Han, Z. 2018, SPPMA, 48, 079803
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Dufon, S., Cunha, K., Becker, S. R., & Smith, V. V. 2001, ApJ, 552, 309
Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, RAA, 12, 755
Duffon, P. L., Smartt, S. J., & Hamblly, N. C. 1999, A&AS, 139, 231
Esteban, C., Fang, X., García-Rojas, J., & Toribio San Cipriano, L. 2017, MNRAS, 471, 987
Fitzpatrick, E. L., & Garmany, C. D. 1990, ApJ, 363, 119
Grevesse, N., & Sauval, A. J. 1998, SSRv, 85, 161
Guo, Y., Li, J., Xiong, J., et al. 2021, arXiv:2109.09775
Han, Z.-W., Ge, H.-W., Chen, X.-F., & Chen, H.-L. 2020, RAA, 20, 161
Heber, U. 2009, ARA&A, 47, 211
Herbig, G. H. 1995, ARA&A, 33, 19
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Hunter, I., Brott, I., Langer, N., et al. 2009, A&A, 496, 841
Harley, J. R., Pols, O. R., & Tout, C. A. 2000, MNRAS, 315, 543
Langer, N., & Maeder, A. 1995, A&A, 295, 685
Langer, N., Schürmann, C., Stoll, K., et al. 2020, A&A, 638, A39
Lanz, T., & Hubeny, I. 2003, ApJS, 146, 417
Lanz, T., & Hubeny, I. 2007, ApJS, 169, 83
Li, J., Liu, C., Zhang, B., & Tian, H. 2021, ApJS, 253, 45
Liu, C., Fu, J., Shi, J., et al. 2020, arXiv:2005.07210
Liu, Z., Cui, W., Liu, C., et al. 2019, ApJS, 241, 32
Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, ApJ, 15, 1095
Maíz Apellániz, J. 2010, A&A, 518, A1
McErlean, N. D., Lennon, D. J., & Dufon, P. L. 1999, A&A, 349, 553
McEvoy, C. M., Dufon, P. L., Smoker, J. V., et al. 2017, ApJ, 842, 32
Morgan, W. W., & Keenan, P. C. 1973, ARA&A, 11, 29
Morton, D. C., & Adams, T. F. 1968, ApJ, 151, 611
Ness, M., Hogg, D. W., Rix, H. W., Ho, A. Y. Q., & Zasowski, G. 2015, ApJ, 808, 16
Nieva, M. F., & Przybilla, N. 2012, A&A, 539, A143
Ojala, M., & Garriga, G. C. 2009, J. Mach. Learn. Res., 11, 1833
Panagia, N. 1973, AJ, 78, 929

11
Sadowski, A., Belczynski, K., Bulik, T., et al. 2008, ApJ, 676, 1162
Sana, H. 2017, in IAU Symp. 329, The Lives and Death-Throes of Massive Stars, ed. J. J. Eldridge et al. (Cambridge: Cambridge Univ. Press), 110
Simón-Díaz, S., Castro, N., Garcia, M., Herrero, A., & Markova, N. 2011, BSRSL, 80, 514
Trundle, C., Dufton, P. L., Hunter, I., et al. 2007, A&A, 471, 625
Urbaneja, M. A., Herrero, A., Bresolin, F., et al. 2005, ApJ, 622, 862
Xiang, M.-S., Rix, H.-W., Ting, Y.-S., et al. 2021, ApJS, 253, 22
Yoon, S. C., Woosley, S. E., & Langer, N. 2010, ApJ, 725, 940
Zhang, B., Li, J., Yang, F., et al. 2021, ApJS, 256, 14
Zhang, B., Liu, C., & Deng, L.-C. 2020a, ApJS, 246, 9
Zhang, B., Liu, C., Li, C.-Q., et al. 2020b, RAA, 20, 051
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723

The Astrophysical Journal Supplement Series, 257:54 (12pp), 2021 December
Guo et al.