Astroinformatics
Proceedings IAU Symposium No. 325, 2016
M. Brescia, S.G. Djorgovski, E. Feigelson,
G. Longo & S. Cavuoti, eds.

doi:10.1017/S1743921316013193

An Automated Galaxy Spectra Recognition Method Basing on Spectral Lines Information

Jiannan Zhang[1], **Xiaoyan Chen**[1], **Yue Wu**[1] **and Xiangru Li**[2]

[1]National Astronomical Observatory, Chinese Academy of Sciences, 100012, Beijing, China
email: jnzhang@bao.ac.cn Chenxy@bao.ac.cn Wuyue@bao.ac.cn

[2]School of Mathematical Sciences, South China Normal University 510631, Guangzhou, China
email: xiangru.li@qq.com

Abstract. For the vast amounts of spectra produced by LAMOST, the pipeline basing on PCAZ method is limited by the bad flux calibration and low S/N data. This work focuses on the study of the efficient recognition methods of galaxy spectra of LAMOST basing on spectral lines information. The new method searches spectral lines and extracts the information of spectral lines (position, height, and width *et al.*) automatically. Using the spectral lines information which are less influenced by the quality of flux calibration and the S/N ratio, galaxy spectra are recognized with the redshift measured through spectral lines matching method. The experiment verified it is feasible for the LAMOST galaxy spectra: the correct recognition rate $> 80\%$ for the data with $SNR\_g > 5$, and $> 90\%$ for the data with $SNR\_r > 5$. Compared with the redshift of SDSS, the systematic error of our method is 0, and the standard deviation of the error is 0.0002.

Keywords. methods: data analysis, techniques: spectroscopic, galaxies: fundamental parameters (classification, colors, luminosities, masses, radii, etc.), telescopes:LAMOST.

1. Introduction

The LAMOST survey project has released more than 4130,000 DR2 spectra data. The majority of the released galaxy spectra are recognized through eyecheck. There are three reasons for the limitation of LAMOST pipeline on galaxy spectra recognition: bad flux calibration, low SNR of extra-galaxy spectra and the PCAZ method applied by the pipeline. The classical PCAZ method relies on the good flux calibration for the examples of SDSS pipeline (Bolton 2012) and GAMA(Baldry 2014). It is necessary to apply new processing methods to improve the performance of the extra-galactic spectra recognition and redshift measurement. This paper presents a galaxy recognition and measurement system (named as galaxy module)which searches spectral lines and extracts the information of spectral lines (position, height, and width *et al.*) automatically. Using the spectral lines information which are less influenced by the quality of flux calibration and the S/N ratio, galaxy spectra are recognized, meanwhile the redshifts are measured through spectral lines position.

2. The galaxy recognition and measurement system

2.1. *Galaxy module*

The galaxy module consists three units: continuum determination, spectral lines detection and measurement, spectral lines matching and redshift determination. To extract the spectral lines, the first step is fitting the spectral continuum. Researchers have explored several methods to fitting the spectral continuum. Lee (2008) developed high order polynomial repeated fitting method to fit the stellar continuum spectrum of SDSS. Wu (2011)applied high order polynomial to normalizing the LAMOST stellar spectra for the stellar parametrization. We have developed a median filter method to fitting the continuum, which applies to any type of spectra. For the spectral lines detection unit, we defined a continuous threshold curve along the spectral flux to detect lines flux points which out of the threshold curve. Once all the spectral lines points are determined, the lines information such as centers, heights and widths are measured through a gaussian function fitted to each group of lines points. At last, if the most of the spectral lines centers are cross matched with galaxy spectral lines list in success, the spectrum is classified as galaxy and its redshift is determined by the lines redshifts. The detail of the continuum fitting method, spectral lines detection method and spectral lines matching method are described in the following galaxy module procedure.

2.2. *The procedure of the galaxy module*

The whole procedure of the galaxy module is as follows.

Noise processing. A Gaussian filter with sigma of 1.5 times of wavelength step is applied to the spectrum to eliminate noise.

Continuum extraction. Spectral continuum is extracted with median filters. Firstly a median filter of 60 pixels width smoothes the spectrum; Then a median filter of 300 pixels width smoothes the processed spectrum above to obtain the final continuum. The normalized spectrum is achieved through original spectrum minus final continuum.

Spectral lines detection. Search all the lines points that the flux point above or below of the threshold curve which is determined by the 3σ of local flux of the normalized spectrum.

Spectral lines measurement. Search all the lines peak points and the lines wing edge points, then fit each group of the line points with a Gaussian function to determine the line center, width and height.

Spectral lines matching. Match all the lines centers with the galaxy lines list and compute all the lines redshifts. If most of the galaxy lines list are matched in success, the spectrum is set to be galaxy, and the corresponding lines redshifts are the raw redshifts of the spectrum.

Redshift. Average the raw lines redshifts to obtain the final redshift of the spectrum.

Fig. 1 illustrates the procedure of spectral lines detection and measurement for a galaxy spectrum.

3. The performance of galaxy spectra recognition and measurement module

3.1. *The correct ratio of galaxy recognition*

The test data are from the LAMOST spectra observed during March of 2014, namely between 1 st. and 26 th. of March. According to the result of the LAMOST pipeline classification, the spectra that classified as Unknown, GALAXY, QSO, STAR with confidence less than 0.80 and the spectra that the target type is extra-galaxy source, are selected to form the test data set, which are 59614 in all. There are 2155 spectra are affirmed to be galaxy spectra after eye-check of all the test data. The galaxy module recognized 1991 galaxy spectra correctly. Fig. 2 analyzes the correct ratio of the result versus SNR. For the spectra with $SNR\_g > 5$, the correct recognition ratio is more than

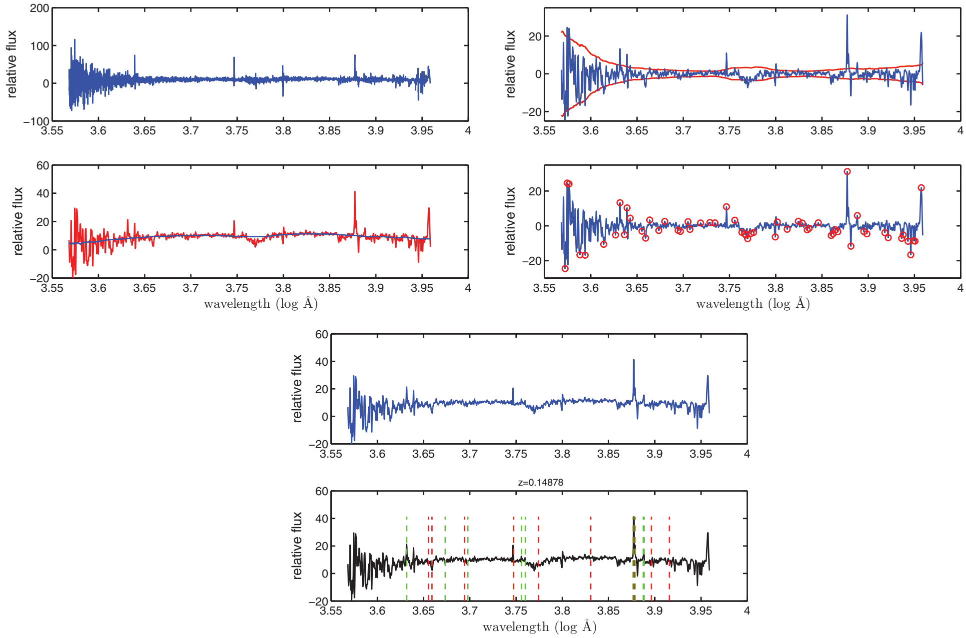

Figure 1. The procedure of spectral lines detection and measurement for an emitting line galaxy spectrum. Upper left: the original spectra, denoised spectrum and the extracted continuum. Upper right: the spectral lines detection through threshold curves (red curve), and the peak points of the detected spectral lines marked with circles. Down: identified galaxy spectral lines through lines matching method, where absorption lines are marked by red dash lines and emmiting lines are marked by green dash lines.

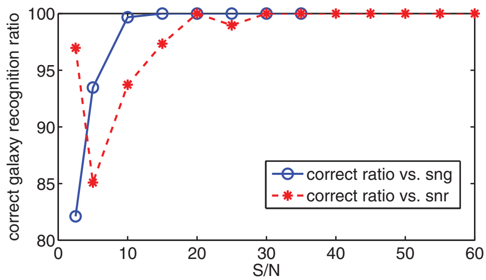

Figure 2. Correct ratio of galaxy classification VS. SNR. Red line: correct ratio with SNR_g; Blue line: correct ratio with SNR_r.

90%, while for the spectra with $SNR\_r > 5$, the correct recognition ratio is more than 80%.

3.2. *The accuracy of redshift measurement*

Among the 1991 galaxy spectra that galaxy module recognized correctly, 1457 galaxy spectra have spectra data in SDSS DR12 catalog through cross matching method. Fig. 3 compares the redshifts of the 1457 LAMOST spectra recognized and measured by galaxy module and the redshifts from corresponding SDSS spectra. Fig. 4 demonstrates the histogram of the differences between redshifts of LAMOST spectra by galaxy module and the redshifts from the corresponding SDSS DR12 catalog. Fitting of Gaussian function for the histogram shows that the systematic difference is 0 and the standard deviation of the difference is 0.0002 (about 60 km/s).

https://doi.org/10.1017/S1743921316013193 Published online by Cambridge University Press

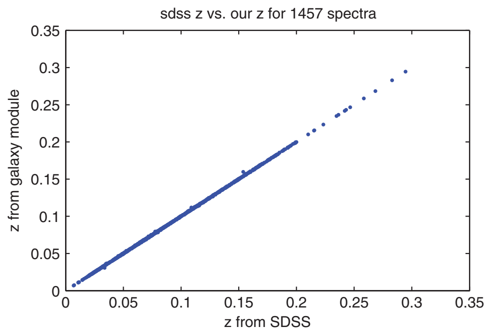

Figure 3. Comparison between the redshifts of 1457 LAMOST galaxy spectra recognized and measurement by galaxy module and the redshifts of corresponding SDSS galaxy spectra.

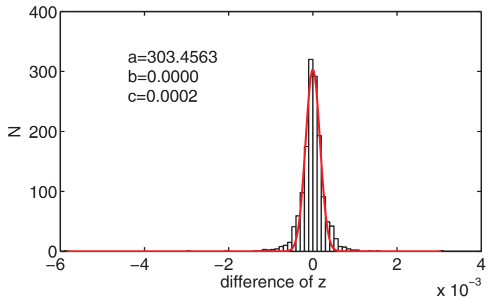

Figure 4. The histogram of the differences between redshifts of LAMOST spectra by the galaxy module and the redshifts from the corresponding SDSS DR12 catalog.

4. Summary

We have developed a galaxy spectra recognition and redshift measurement system for the low-resolution LAMOST spectra. The system is composed of three units: continuum fitting, spectral lines detection and measurement, spectral lines matching and redshift determination. The experiment result shows its good correct galaxy recognition ratio and the accurate redshift measurement for the spectra with $SNR\_g > 5$ or $SNR\_r > 5$. The galaxy module basing on spectral lines method will be applied in the future LAMOST data processing procedure.

Acknowledgements

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 Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences. This work is supported by the National Natural Science Foundation of China (No.11403059 and No.61273248), and the Joint Astronomical Fund of NSFC and CAS (No.U1531242).

References

Bolton, S., Schlegel, J., Aubourg, E., Bailey, S., *et al.* 2012, *AJ*, 144, 144
Baldry, K., Alpaslan, M., Bauer, E., Bland-Hawthorn, J., *et al.* 2014, *MNRAS*, 442, 2440
Wu, Y., Luo, A-Li, Li, H., Shi, J., *et al.* 2011, *RAA*, 11, 924
Lee, Y., Beers, C., Sivarani, T., Allende, C., *et al.* 2008, *AJ*, 136, 2022