LAMOST DR5 Spectral Clustering for Stellar Templates Construction

Xiao Kong,[1] and A-Li Luo[1]
[1]*National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China;* kongx@nao.cas.cn

Abstract.
The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) released the fifth spectral data (DR5) this year, containing 8,171,443 star, 153,090 galaxy, 51,133 QSO and 642,178 unknown types, all of which are classified by the LAMOST 1D pipeline. This pipeline is used for spectral analysis, aiming to determine the spectral type and redshifts of the spectra observed by LAMOST by matching them with spectral templates. Generally, the performance of the stellar classification greatly depends on the quality of templates. In this experiment, we build 197 templates that are able to used as stellar templates for pipeline. These templates are supposed to increase the number of types and the accuracy of the classification. In addition, another 19 cluster centers are still remained to be identified.

1. Introduction

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)(Cui et al. 2012), with the ability of capturing 4,000 objects during one exposure, is a special quasi-meridian reflecting Schmidt telescope located in Xinglong Station of national Astronomical Observatory, China(Yao et al. 2012). It has begun to release spectral data since 2012 (the prior data release, PDR). At the beginning of 2018, the LAMOST released its 5th spectral data, DR5, including 9,017,844 spectra (Table 1).

Table 1. The number of released spectra from LAMOST DR of each year.

	DR1	DR2	DR3	DR4	DR5
Star	1,944,329	3,784,461	5,268,687	6,856,896	8,171,443
Galaxy	12,082	37,206	61,815	118,657	153,090
QSO	5,017	8,630	16,351	36,374	51,133
Unknown	243,268	306,185	408,273	652,146	642,178
Total	2,204,696	4,136,482	5,755,126	7,664,073	9,017,844

All the spectral type are assigned by LAMOST 1D pipeline using template matching. The current stellar templates(Wei et al. 2014) were constructed from spectral data of LAMOST DR1(Luo et al. 2015). As shown in Table 2, there are 183 stellar spectral that served as the templates for classification, and they are able to help pipeline software classify the stellar spectra at a high correction ratio.

Table 2. The number of spectra within the current LAMOST stellar template.

Subclass[a]	O	B	A	F	G	K	M	Carbon	CV	DoubleStar	WD	Total
Number	2	2	49	25	24	36	38	3	1	1	2	183

[a] "Subclass" is adopted from the data archive of LAMOST DR5.

With the quantity of stellar spectra photoed by the LAMOST increasing at an amazing rate, however, some more rare objects appear in the huge amount of spectral data sets that cannot be identified by the pipeline using the current templates, e.g. DB white dwarfs. We utilize all the ≈9 million spectra from LAMOST DR5 as the material of new templates for LAMOST 1D pipeline.

2. Clustering

The basic idea of construction is hierarchical clustering: (1) dividing all the spectra into different groups according to their spectral type; (2) clustering spectra within each group into different cluster centers; (3) re-clustering all the centers.

At the beginning, full-spectra template-matching is adopted to assign each spectrum to a certain type (could be any type from all kinds in Table 2, QSO and galaxy) and the redshift (z) can be obtained. Consider the affect of noise, spectra from each type are divided into two groups: signal to noise ratio (S/N) $\leq$ and $>$ 10, respectively. Then 24 groups are built.

Afterward, we move all spectral to the rest frame and set 100 cluster centers for each group befor employing k-means(MacQueen 1967) to these groups respectively. All the 2,400 cluster centers are adopted k-means again and 500 centers are screened. These centers corresponds to different amount of spectra.

Those centers corresponding to spectra less than 1000 are abandoned at first. We visually inspect all of them and discard 181 centers due to low spectral quality or the similarity between different group centers. Furthermore, the types of the remained clustering centers are assigned by the subclass of spectra from LAMOST DR5. Meanwhile, 19 templates whose subclasses are difficult to determine need to be identified in our further work.

Some of the cluster centers are illustrated in the next section.

3. Images

4. References

References

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., & et al. 2012, Research in Astronomy and Astrophysics, 12, 1197

Luo, A.-L., Zhao, Y.-H., Zhao, G., & et al. 2015, Research in Astronomy and Astrophysics, 15, 1095

MacQueen, J. 1967, in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics (Berkeley, Calif.: University of California Press), 281. URL https://projecteuclid.org/euclid.bsmsp/1200512992

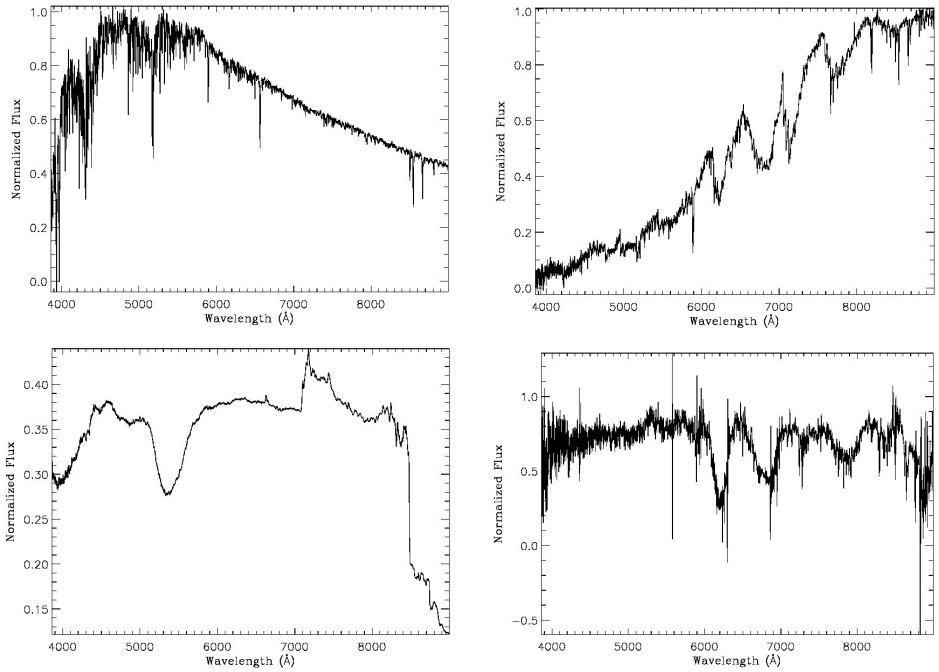

Figure 1. Some cluster centers. The bottom ones are two examples that need to be identified in the next step. *Upper left:* star, G5. *Upper right:* star, M4.

Wei, P., Luo, A., Li, Y., & et al. 2014, AJ, 147, 101
Yao, S., Liu, C., Zhang, H.-T., & et al. 2012, Research in Astronomy and Astrophysics, 12, 772. 1206.3574