ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-201903.00229

Spline-Based Automatic Normalization Method for Stellar Spectra (Postprint)

Authors: Luo Feng, Liu Chao, Zhao Yongheng

Date: 2019-03-25T00:00:00+00:00

Abstract

The observed spectra of stars generally consist of continuum, spectral lines, and noise. The continuum is a smooth, continuous spectrum in which the radiation flux varies with wavelength due to blackbody radiation. Research such as spectral classification and estimation of stellar physical parameters all rely on accurate extraction of continuum and spectral line information. Therefore, the primary task in spectral data processing is to fit the continuum and extract spectral line features by normalizing the spectrum. Current continuum fitting methods mainly include polynomial fitting, median filtering, wavelet filtering, among others. Existing methods exhibit varying degrees of limitations in cases of low signal-to-noise ratio, cosmic ray interference, and presence of emission lines, primarily in terms of robustness and accuracy. Currently, no automated method has been applied to the normalization of LAMOST' s 10^7 spectra. In this era of astronomical data avalanche, it is extremely urgent to research and develop a stellar spectral normalization algorithm that is applicable to a wider range of temperatures, signal-to-noise ratios, and wavelength coverage, with better universality and capable of automated processing. Based on careful analysis of different types of spectra, we propose a continuum fitting method based on fixed window partitioning. This method screens and extracts data points that can reflect continuum characteristics in the spectrum, and produces a more accurate continuum by finely controlling the smoothness of the spline function. Experiments using spectra from LAMOST with different spectral types, temperature ranges, and wavelength coverage demonstrate that the algorithm proposed in this paper possesses good precision and universality.

Full Text

Preamble

Automatic Normalization Method for Stellar Spectra Based on Spline Functions

Luo Feng[1],[2], Liu Chao[1],[2], Zhao Yongheng[1],[2]
(1. National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101;
2. University of Chinese Academy of Sciences, Beijing 100101)

Abstract

The observed spectrum of a star generally consists of three components: continuum, spectral lines, and noise. The continuum is a smooth, continuous spectrum representing the variation of radiation flux with wavelength due to blackbody radiation. Both spectral classification and estimation of stellar physical parameters depend on accurate extraction of continuum and spectral line information. Therefore, the primary task in spectral data processing is to fit the continuum and extract spectral line features through normalization.

Current continuum fitting methods mainly include polynomial fitting, median filtering, and wavelet filtering. Existing approaches exhibit varying degrees of limitation in cases of low signal-to-noise ratio (SNR), cosmic ray interference, and presence of emission lines, primarily reflected in robustness and accuracy. At present, no automated method has been applied to normalize the 10 spectra from LAMOST. Given the avalanche of astronomical data, research and development of a stellar spectral normalization algorithm with better universality and automated processing capability—applicable across broader temperature ranges, SNR levels, and wavelength coverage—has become extremely urgent.

Based on careful analysis of different spectral types, we propose a continuum fitting method based on fixed-window partitioning. This approach screens and extracts data points that reflect continuum characteristics from the spectrum, producing a more accurate continuum through fine control of spline function smoothness. Experiments using LAMOST spectra covering different spectral types, temperature ranges, and wavelength coverage demonstrate that the proposed algorithm achieves good precision and universality.

Keywords: Continuum normalization; LAMOST; Stellar spectrum; Spline function

0 Introduction

As a major national scientific project, LAMOST has overcome the challenge of simultaneously achieving large field of view and large aperture in astronomical telescopes, becoming the largest wide-field telescope in the world and marking another milestone in China' s optical telescope development. As of June 2018, LAMOST completed seven years of sky surveys, obtaining over ten million spectra for the first time and becoming the world' s first survey project to exceed this magnitude. Unfortunately, no automated method currently exists for normalizing LAMOST' s 10 spectra, and this paper focuses on addressing this critical step.

Each observed spectrum comprises continuum, spectral lines, and noise. Spectral lines enable analysis of stellar chemical composition and, due to the Doppler effect, carry radial velocity information. Therefore, accurate extraction of spectral lines from raw spectra represents an important step in stellar spectral research and data processing, with significant implications for subsequent analyses.

Eliminating the continuum first requires fitting it. Existing continuum fitting methods include polynomial fitting, median filtering, wavelet filtering, and morphological filtering. Semi-automatic methods using astronomical software also exist: first manually selecting appropriate data points and choosing a functional form, then performing polynomial fitting with software to obtain the continuum. Such methods rely on human intervention and cannot meet the real-time processing demands of massive spectral data in current sky survey projects.

The Sloan Digital Sky Survey (SDSS) project's SSPP (SEGUE Stellar Parameter Pipeline) employs a piecewise polynomial fitting approach. The spectrum is divided into blue and red segments. After removing strong Balmer lines, a 9th-order polynomial fits the blue segment while a 4th-order polynomial fits the red segment, discarding data points beyond 3 standard deviations. The two continuum segments are then spliced together and fitted again with a 9th-order polynomial to obtain the final continuum. This method is somewhat cumbersome and computationally intensive.

The method proposed in this paper divides the spectrum into several fixed windows (containing equal numbers of pixels), selects median points within each window, and then applies screening strategies to discard certain reference points. This approach effectively avoids emission lines, absorption lines, and cosmic rays. By controlling the "angle"between reference points, very small smoothness values can be used to construct spline curves, enabling precise continuum fitting.

1 Methodology

The proposed continuum normalization method consists of three main steps: First, the input spectrum is divided into several windows based on the total number of data points, with each window selecting a median flux point to form a reference point set. Second, screening strategies discard potentially problematic data points from this set. Finally, an empirical smoothness value for the spline function is selected based on the wavelength position of the maximum flux point in the reference set, and the continuum is fitted.

1.1 Window Partitioning and Reference Point Set Construction

To ensure sufficient data points for fitting, we adopt the following windowing strategy based on the total number of data points n in the input spectrum. Let w represent the number of pixels per window (window width):

After determining the window width from the table above, the spectrum is

partitioned accordingly. We then traverse all windows, selecting the pixel whose flux value equals the median of all points in that window as a reference point and adding it to the reference point set Srp.

1.2 Reference Point Screening

The reference point set Srp constructed in the previous step only reflects the general shape of the spectrum, with some points potentially located on absorption or emission lines, as shown in the figures:

[Figure 1: see original paper]
[Figure 2: see original paper]

Theoretically, the continuum should closely approximate a blackbody radiation spectrum, which can be divided into two monotonic intervals: monotonically increasing on the left side of the peak and monotonically decreasing on the right side. We therefore desire that reference points used to construct the fitting curve reflect this characteristic. To achieve this goal, we implement a two-step screening process.

(1) Discarding Local Minimum Points

For a reference point with index i in Srp and flux $f(i)$, if i satisfies:

$$f(i-1) > f(i) \text{ and } f(i) < f(i+1), \quad i \geq 1$$

then reference point i is removed from Srp. Each pass begins at i=1 and traverses all reference points in Srp. After one pass, some points may become new local minima, necessitating additional passes. Let t represent the number of passes. We must ensure two conditions: for spectra with normal continua, the number of elements in Srp no longer decreases within t passes; for spectra with abnormal continua, sufficient reference points remain after t passes to construct the fitting curve. Experiments show that an appropriate value is:

$$t = \text{len}(S_{rp})//30 + 2$$

After this processing, the reference point sets for the spectra in Figures 1 and 2 become:

[Figure 3: see original paper]
[Figure 4: see original paper]

(2) Discarding Points with Large "Angles"

To obtain more precise continuum fitting, we require small smoothness values for spline construction. However, if the angle formed between a reference point and its neighboring points is too large, it can cause significant deviation in the fitted curve. Such reference points should be removed from Srp, as illustrated in Figure 5.

[Figure 5: see original paper]
[Figure 6: see original paper]

As shown in Figure 6, for a point i in Srp with wavelength $w(i)$ and flux $f(i)$, let (i) be the angle between its connections to left and right points, with triangle side lengths a, b, and c. Geometrically:

$$a = \sqrt{(w(i) - w(i-1))^2 + (f(i) - f(i-1))^2}$$
$$b = \sqrt{(w(i+1) - w(i))^2 + (f(i+1) - f(i))^2}$$
$$c = \sqrt{(w(i+1) - w(i-1))^2 + (f(i+1) - f(i-1))^2}$$

The angle (i) can be calculated using the cosine theorem. However, flux values do not correspond to length units, making it difficult to unify the units of w and f. Moreover, since flux in our spectral data is uncalibrated relative flux, the angle calculated directly using the cosine theorem varies with the flux value range and necessary multiplication/division operations during spectral processing. We therefore implement the following approach:

For a reference point i in Srp, let the wavelength difference with the left point be $wd-$ and with the right point be $wd+$:

$$wd_{-} = w(i) - w(i-1)$$
$$wd_{+} = w(i+1) - w(i)$$

Similarly for flux values:

$$fd_{-} = f(i) - f(i-1)$$
$$fd_{+} = f(i+1) - f(i)$$

We introduce scaling factors rw and rf:

$$r_{w} = \frac{wd_{-}}{wd_{+}}$$
$$r_{f} = \frac{fd_{-}}{fd_{+}}$$

Finally, we compute the criterion $r = |r\text{w - } r\text{f}|$ and introduce a threshold $r0$. If $r > r0$, point i is considered to have too large an angle and will interfere with precise fitting, and should be removed from Srp. Through experimentation, we find an effective value of $r0 = 0.5$.

Using ratios eliminates dimensional dependence, making rw and rf invariant regardless of the units used for w and f. This provides a more reliable basis

for identifying problematic reference points. The result after processing the spectrum in Figure 5 is shown in Figure 7.

[Figure 7: see original paper]

1.3 Smoothness Calculation

Fitting quality depends not only on reference point selection but also on appropriate smoothness values. For the same spectrum with identical reference points, different smoothness values produce different results, as shown in Figure 8.

[Figure 8: see original paper]

Smaller smoothness better captures detailed undulations in the spectral profile, but excessively small values cause frequent oscillations, preventing the fitted curve from representing the true continuum. Larger smoothness eliminates violent oscillations, but excessive values make the fitted curve too flat, also poorly representing the continuum. Notably, for a given smoothness value, spline curves exhibit different behaviors across different flux intervals. Before fitting, we normalize all spectral data by their maximum values: $flux = flux / \max(flux)$.

Experimental analysis shows that higher-temperature spectra require smaller smoothness to accommodate sharper turns in the blue end, while lower-temperature spectra require larger smoothness to dampen upward and downward trends in reference points. A spectrum' s temperature is reflected in the continuum by the position of its flux peak. We select the maximum flux point in the reference point set and use its wavelength position to estimate the temperature range and select the corresponding smoothness $s0$. To accommodate local normalization scenarios with fewer reference points, we introduce a scaling coefficient c (default value 1) to adjust the final smoothness. Specific values are provided in Tables 2 and 3.

The final smoothness used in spline construction is:

$$s = c \times s_0$$

2 Experimental Results

2.1 Processing Results Under Various Conditions

To evaluate the proposed method, we randomly selected 10,000 spectra of different types from LAMOST. Representative spectra at different temperatures were chosen for fitting, as shown in Figure 9. The left panel displays the original spectrum (blue solid line) and fitted curve (red solid line); the right panel shows the corresponding normalized spectrum (green solid line). The results demonstrate that our method achieves good fitting performance across different temperatures.

[Figure 9: see original paper]

As shown in Figure 10, the left panel shows the original spectrum (magenta solid line) and spline fitting curve (green solid line), while the right panel shows the normalized spectrum (indigo solid line). The method works well with different wavelength coverage ranges, enabling automatic applicability to local normalization scenarios without manual intervention.

[Figure 10: see original paper]

Figure 11 demonstrates that for different SNR levels, the fitting effect remains largely unaffected. The results are very close to those achieved by the normalization method specifically developed for low-quality spectra in reference [9].

[Figure 11: see original paper]

As shown in Figure 12, the method effectively identifies and accurately processes spectra containing emission lines and cosmic rays.

[Figure 12: see original paper]

2.2 Error Analysis and Applicability Range

Due to instrumental effects, LAMOST spectra commonly exhibit three types of instrumental deformation patterns, shown in the left subplot of Figure 8. To test whether our method can effectively handle these deformations, we randomly selected nine spectra with effective temperatures between 4000–8000 K. Each normalized spectrum was multiplied by these three deformation curves and re-normalized. The histogram in the right subplot of Figure 8 shows the distribution of residuals between the two normalization results, demonstrating that our algorithm' s error averages on the order of 10^{3} with dispersion on the order of 10^{2}.

Experimental analysis indicates the algorithm's applicability range as: $5 < \mathrm{SNR} < 600$, wavelength coverage 3700–9000 Å, and effective temperature range 3000–50000 K. For SNR below 5, the algorithm becomes unstable with decreasing overall accuracy. Spectra with effective temperatures below 3000 K also yield unstable and often unreliable results. Spectra with SNR above 600, wavelength ranges below 3700 Å or above 9000 Å, and effective temperatures above 50000 K were not tested. Additionally, the algorithm is suitable for medium- to low-dispersion spectra; high-dispersion spectra require parameter fine-tuning.

We note that the parameter values (w, t, $r0$, $s0$, c) were selected to accommodate instrument characteristics. For spectra from other sources, the method remains applicable but requires slight adjustment of initial parameter values.

3 Conclusions and Outlook

Normalization is a crucial step in spectral data processing, and its accuracy directly affects the correctness and precision of subsequent analyses. Building

upon domestic and international research, we propose a spline function fitting method based on fixed-window partitioning, with empirical parameter values refined through extensive experimentation. Validation using additional data demonstrates that our method achieves automated processing while broadening applicability, offering superior universality compared to other methods and unique advantages for large-scale automated processing scenarios.

Various stellar spectral normalization methods exist, each with strengths and weaknesses in different contexts. Our algorithm focuses on enabling automated processing and universality expansion for LAMOST' s massive spectral dataset. All involved parameter values were empirically derived from extensive experiments. However, an important unresolved issue during algorithm development is the lack of a strict optimal solution evaluation criterion to assess normalization quality, limiting the algorithm to basic error handling such as preventing negative continuum values. If a quantifiable and computationally implementable evaluation standard could be established, processing results could be assessed and parameters automatically adapted for reprocessing until optimal results are achieved.

With the explosive growth of astronomical data, more efficient and intelligent normalization methods and data processing programs are needed. Future work will focus on spectral normalization for cool stars and special stars such as carbon stars.

References

[1] Zhao Yongheng. LAMOST celestial spectral survey [J]. Physics, 2015, 44(04): 205-212.
[2] Automatic Normalization and Equivalent-Width Measurement of High-Resolution Stellar Spectra [J]. Chinese Journal of Astronomy and Astrophysics, 2006(06): 689-696.
[3] Starck J L, Siebenmorgen R, Gredel R. The Astrophysical Journal, 1997, 482: 1011.
[4] Zhao Ruizhen, Luo Ali. A Novel Method for Continuum Normalization of Astronomical Spectrum Signals [J]. Spectroscopy and Spectral Analysis, 2006(03): 587-590.
[5] Lee Y S, Beers T C, Sivarani T, et al. The SEGUE stellar parameter pipeline. I. Description and comparison of individual methods [J]. The Astronomical Journal, 2008, 136(5): 2022.
[6] Zeng Jinyan. Quantum Mechanics: Volume I [M]. Science Press, 2013: 3-4.
[7] Yu Jingjing, Pan Jingchang, Meng Fanlong, Wei Peng. Automatic detection of continuum anomalies in LAMOST spectra based on distance metric [J]. Spectroscopy and Spectral Analysis, 2017, 37(07): 2246-2249.
[8] Schumaker L. Spline functions: basic theory [M]. Cambridge University Press, 2007.
[9] Wu Minglei, Pan Jingchang, Yi Zhenping, Wei Peng. A method to fit low-quality stellar spectrum [J]. Spectroscopy and Spectral Analysis, 2018,

38(03): 963-967.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.