

一种重复观测低信噪比光谱的处理方法

刘园园¹, 陈建军², 邱波^{1*}, 范晓东¹, 魏诗雅¹, 宋涛¹, 段福庆^{3*}

1. 河北工业大学电子信息工程学院, 天津 300401

2. 中国科学院国家天文台, 北京 100012

3. 北京师范大学, 北京 100875

摘要 目前我国LAMOST光谱巡天已发布超过760万条的天体光谱,对其中大量的低信噪比光谱的处理一直是业内公认的难题。针对天体光谱中重复观测的光谱,提出了一种新的处理方法。该方法的主要内容为:对每一组重复观测光谱,选择其红移值的差距在一定范围内的组别,然后使用一种基于信噪比加权的最优叠加方法来提高光谱的信噪比。通过对LAMOST DR4中所有重复观测光谱进行处理,证明该方法对于提高低信噪比重复观测光谱的信噪比十分有效。使7571组恒星光谱的信噪比达到参数测量的标准,3357组类星体和星系光谱的信噪比得到提高,平均提高率为56.38%;并且获得了43021个双星候选体。

关键词 重复观测光谱;低信噪比;CCD噪声源;双星候选体

中图分类号: P157.2 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2018)07-2311-04

引言

我国的大天区面积多目标光纤光谱望远镜LAMOST (large sky area multi-object fiber spectroscopic telescope),超过了国际上目前已完成的或正在进行中的各类大视场多天体光纤光谱巡天计划,是当今世界上天体光谱获取率最高的望远镜,可实现短时间内对多个天体进行重复观测^[1]。

目前,在LAMOST DR4总星表中有超过760万的天体光谱,其中包括680多万条恒星光谱。AFGK星表中的430多万条恒星光谱,通过参数测量可以得到恒星的大气参数,包括有效温度 T_{eff} ,恒星表面重力 $\log g$,金属丰度 $[\text{Fe}/\text{H}]$ 以及日心视向速度^[2]。这些参数对于研究银河系结构和恒星物理有重要的意义^[3]。但是,由于望远镜系统在观测目标星体时受到大气层扰动、光的衍射以及环境和检测噪声的影响^[4],使光谱质量下降,一些天体光谱由于信噪比较低而无法进行参数测量。

提出了对重复观测光谱的处理方法,针对红移值的差距在一定阈值范围内的低信噪比的重复观测光谱,通过一种基于信噪比加权的最优叠加方法,实现对低信噪比光谱增强的处理。其中,在给出一组重复观测光谱中的每条光谱分配权重

时,根据其各自信噪比影响因子的不同进行分配。并且在此基础上对影响因子的定义做了改进,减小了信噪比极值对叠加效果的不利影响。本文的贡献是,通过对LAMOST DR4中所有的光谱进行增强处理,共有7571组恒星光谱的信噪比达到参数测量的标准,3357组类星体和星系光谱的信噪比得到提高,平均提高率为56.38%;另外,通过将红移值差距较大的重复观测光谱,定义为双星候选体,共获得了43021个双星候选体。

1 CCD噪声源与信噪比

CCD相机系统中的噪声来源主要有三个:光子噪声、暗电流噪声和读出噪声。在CCD相机曝光时,被测量目标在CCD焦面上产生信号的同时将引入光子噪声,由于CCD像元搜集的光电子服从泊松分布,因此光子噪声强度等于信号强度的平方根。与光子噪声相似,暗电流噪声也表现为泊松分布,它等于在曝光时间内产生的热电子的均方根。天文CCD相机一般工作在致冷状态下,目前已经达到在一定曝光时间内暗电流噪声可以被忽略的温度。读出噪声主要来源于片上的预放大器,该噪声被均匀添加到图像的每个像素^[5]。

在CCD相机系统中,信噪比在广义上是指一个像素点

收稿日期:2017-08-14,修订日期:2017-12-29

基金项目:国家自然科学基金委员会-中国科学院天文联合基金项目(U1631131, U1531242),国家自然科学基金项目(11403059)和河北省科技支撑计划(15212105D)资助

作者简介:刘园园,1992年生,河北工业大学电子信息工程学院硕士研究生 e-mail: 1044386151@qq.com

* 通讯联系人 e-mail: qiubo@hebut.edu.cn; fqduan@bnu.edu.cn

上相应的信号的大小与其自身不确定性的比值,在狭义上是指一个像素点上测量的信号与全部的测量产生的噪声的比值^[6]。由以上对噪声的分析可知,CCD相机系统信噪比的计算公式为

$$\text{SNR} = \frac{PQ_e t}{\sqrt{(P+B)Q_e t + I_d t + N_r^2}} \quad (1)$$

式(1)中, P 表示入射光子流,单位为 photons/pixel/second; B 为背景光子流,与 P 有相同的单位;背景光子流的来源受很多因素的影响,通常为杂散光。假定信号是光的唯一来源, B 可以认为是 0; Q_e 表示 CCD 的量子效率; t 为积分时间,单位为 second; I_d 表示暗电流的大小,单位与 P 相同; N_r 表示读出噪声,单位为电子数均方根/像素。

2 流量和误差的传递

LAMOST 能够同时获得大量不同天体的光谱数据,目标天体光谱通过光谱仪经过光纤传输,最终从 CCD 相机中得到二维光谱图像。由于采用了液氮冷却的方法,CCD 工作在致冷状态下暗电流噪声可以被忽略^[7]。所以在二维光谱图像中,误差的计算公式可以表示为

$$e = \sqrt{f + N_r^2} \quad (2)$$

式(2)中, f 为流量, N_r 为读出噪声。

二维光谱通过预处理模块后,经过光谱能量抽取、波长定标、光纤效率改正、减天光、流量定标、光谱合并和信噪比计算,最终得到一维光谱数据^[8]。这个过程中包括了流量值的运算;当流量进行加减运算时,误差均为各误差的平方和的开方;当流量乘除某个数值时,误差也同样乘除该数值^[9]。因为光谱合并时,流量是由样条函数拟合出来的,所以其误差计算方式如下:

假设原波长点为 $\lambda_1, \dots, \lambda_n$, 新采样点 λ'_m 的误差计算要考虑波长在 $\left[\frac{\lambda'_{m-1} + \lambda'_m}{2}, \frac{\lambda'_m + \lambda'_{m+1}}{2} \right]$ 范围内的 k 个点, e_1, \dots, e_k 为这 k 个点分别对应的误差,假设新采样点的流量是这些点的算术平均,那么误差 e 的计算公式为式(3)

$$e = \frac{\sqrt{e_1^2 + \dots + e_k^2}}{k} \quad (3)$$

天文光谱数据库中,通过计算得到的最终的流量 f 和误差 e 储存在 FITS 文件中, FITS 文件主数据阵列的第一行是流量 f , 第二行误差用 invvar 表示^[10], 其值如式(4)

$$\text{invvar} = \frac{1}{e^2} \quad (4)$$

所以,通过 FITS 文件求光谱信噪比的公式如式(5)

$$\text{SNR} = \frac{f}{e} = f \sqrt{\text{invvar}} \quad (5)$$

3 低信噪比光谱处理方法和实验

方法主要分成三步:首先,筛选出重复观测的光谱并对其分类;然后对重复观测光谱中红移值的差距在一定阈值范围内的低信噪比光谱,通过基于信噪比加权的最优叠加实现

提高光谱信噪比的处理;最后,对红移值差距较大的重复观测光谱进行处理和归类,设为双星候选体。

3.1 光谱的筛选与分类

对 LAMOST 已发布的 DR4 总星表中的光谱数据进行了筛选和分类,主要分为四个步骤实现。

(1) 根据 FITS 头文件,将光谱分为两部分,一类是恒星光谱,另一类是类星体和星系,同时将没有测量出红移值的光谱数据剔除。对于正常的类星体和星系,它们属于河外天体,一般距离地球比较远,其观测光谱信噪比较低,所以对所有重复观测的类星体和星系光谱进行叠加处理,使其信噪比尽可能的提高。

(2) 根据观测日期,将暗月夜和亮月夜的恒星光谱分开。因为恒星在参数测量时,根据暗月夜和亮月夜,对光谱信噪比的要求也不同。暗月夜时, g 波段信噪比小于 6 的不进行参数测量;亮月夜时, g 波段信噪比小于 15 的不进行参数测量。

(3) 分别在暗月夜恒星光谱、亮月夜恒星光谱、类星体和星系光谱这三类数据中寻找重复观测的光谱。根据球面距离公式,当两目标的球面距离小于 2 角秒时,可以被认为是相对同源的多次观测光谱。

(4) 计算每组重复观测光谱的最大红移差 ΔZ 。一组重复观测的光谱之间的红移值存在差异,可能是因为它们的波长定标精度不同,也可能是因为目标天体本身是一个双星。如果不考虑 ΔZ ,直接将光谱叠加,会使谱线被展宽,影响处理效果。根据经验,将 ΔZ 定在小于 5×10^{-4} 的范围,在该范围内叠加产生的误差在可接受范围内,并且将大于该范围的重复观测组作为双星候选体。

3.2 基于信噪比加权的最优叠加

采用一种基于信噪比加权的最优叠加方法,对以上介绍的重复观测的低信噪比光谱进行叠加处理,该方法主要包括光谱调整和优化影响因子两部分。

3.2.1 光谱调整

为了提高信噪比,也为了更好地进行叠加处理,将重复观测的光谱调整到统一尺度,该调整主要分为两个步骤实现。

(1) 在一至六阶多项式拟合中,根据拟合精度自动选取拟合阶数,提取连续谱。

(2) 光谱归一化后将归一化的光谱去乘以该组重复观测光谱中整体流量值较高的连续谱,从而将光谱调整为统一尺度。

实验中利用上述光谱调整方法对一组亮月夜重复观测 5 次的恒星的光谱进行处理,图 1 为连续谱的提取,图 2 表示经过光谱调整,5 条光谱达到了统一尺度。

3.2.2 优化影响因子

在一组重复观测的光谱中,每条光谱在相同波长点上的信噪比往往各不相同。这时如果采用传统的叠加方法,对具有不同信噪比的各点,在叠加时分配相同的权重,显然是不合理的。而是应该根据各自的信噪比值分配不同的影响因子。信噪比较大的波长点分配较大的影响因子,提高其在叠加中的贡献;信噪比较小的波长点分配较小的影响因子,减

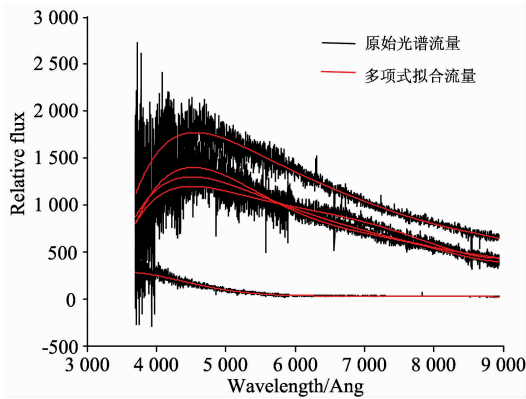


图 1 拟合连续谱

Fig. 1 Continuous spectra fitting

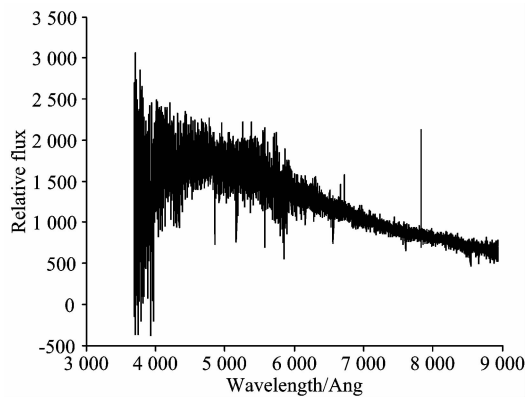


图 2 光谱调整

Fig. 2 Rescaled spectra

小其在叠加中的贡献。由该理论, 信噪比的公式可设计为

$$SNR = \frac{\sum_{i=1}^N (A_i \times f_i)}{\sqrt{\sum_{i=1}^N (A_i \times e_i^2)}} \quad (6)$$

式(6)中, A_i 为一组重复观测光谱中, 第 i 条光谱对总体的影响因子, 作为叠加的权值; f_i 为其流量值; e_i 为其误差; N 为重复观测次数。

通常情况下, 计算影响因子的公式可以表示为

$$A_i = \frac{S_i}{1/N \times \sum_{m=1}^N S_m} \quad (7)$$

式(7)中, S_i 为一组重复观测光谱中, 第 i 条光谱的信噪比; N 为重复观测次数; 分母部分表示同一个波长点处 N 次重复观测的平均信噪比, 该平均信噪比为算术平均。算术平均容易受到极端值的影响。如果一组信噪比值中存在极端值, 即信噪比很高或者很低, 那么信噪比的算术平均对整体信噪比平均水平的代表性将变得很差。因此, 对影响因子 A_i 改进为如式(8)所示

$$A_i = \frac{S_i}{\sqrt{\prod_{n=1}^N S_n}} \quad (8)$$

式(8)中, S_i 仍为一组重复观测光谱中, 第 i 条光谱的信噪比, 而分母由原来的算术平均信噪比变成了受极端值影响较小的几何平均信噪比。

使用该方法对上述图 2 中的光谱数据继续处理, 得到叠加后的光谱图, 如图 3 所示。

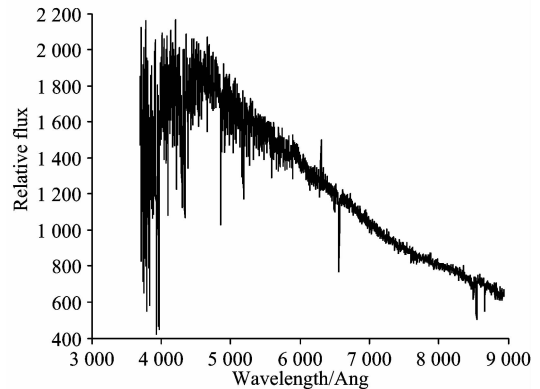


图 3 叠加后光谱

Fig. 3 Spectra stacking

将图 3 中显示的光谱与图 1 中各光谱进行比较发现, 光谱有了明显的改善。处理前, 5 条光谱 g 波段的信噪比分别为 14.18, 14.41, 7.76, 8.73, 6.39, 没有达到参数测量的标准(15); 经过本方法处理后得到的光谱的 g 波段信噪比为 21.28, 该组光谱得到明显改善, 可进行参数测量。

4 结果分析

经过对 LAMOST DR4 中所有的光谱数据做筛选和分类, 红移值的差距 ΔZ 小于 5×10^{-4} 的低信噪比的重复观测光谱共对应 152 472 个重复观测组。按本方法对这些低信噪比光谱做处理, 结果如表 1 所示。

表 1 光谱处理结果

Table 1 Results of spectral processing		
光谱组类型	暗月夜恒星	亮月夜恒星
初始光谱组数	24 703	120 434
改善光谱组数	2 484	5 087
光谱组类型	类星体和星系	
初始光谱组数	7 335	
信噪比提高组数	3 357	
平均提高率/%	56.38	

此外, 将红移值的差距 ΔZ 大于 5×10^{-4} 的重复观测的光谱定义为双星候选体后, 共得到 43 021 个双星候选体, 可供下一步证实。

5 结论

提出了一种简单有效的重复观测光谱的处理方法。通过对重复观测光谱中红移值差范围的设定, 选定该范围内的低

信噪比重复观测光谱; 然后使用一种基于信噪比加权的最优叠加方法, 在经过光谱的调整后, 通过优化影响因子减小信噪比极值对叠加效果的不利影响, 最终得到了改善后的增强光谱。实验结果表明, 该方法对低信噪比光谱的信噪比有较

高的提高率。此外, 还找出了数量众多的双星候选体, 为今后做进一步的证实准备好了数据。未来将改进叠加算法, 使更多低信噪比光谱得到更好的增强。

References

- [1] ZHAO Yong-heng(赵永恒). *Physics(物理)*, 2015, 44(4): 205.
- [2] REN Juan-juan, LUO A-li, ZHAO Yong-heng(任娟娟, 罗阿理, 赵永恒). *Progress in Astronomy(天文学进展)*, 2014, 32(3): 462.
- [3] DAI Wen(戴 闻). *Physics(物理)*, 2017, 46(2): 90.
- [4] AN Ran, PAN Jing-chang, YI Zhen-ping(安 冉, 潘景昌, 衣振萍). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2017, 37(1): 273.
- [5] Badenes C, Hughes J P, Cassamchenai G. *Astrophysical Journal*, 2012, 680(2): 1149.
- [6] Dubner G, Duvidovich L, Giacani E. *Astrophysical Journal*, 2016, 819(2): 98.
- [7] Bhanage R, Borde S, Joshi K. *Co-Operative Communication with SNR Weighted Algorithm in Cognitive Radios*, 2015 IEEE International Conference on Pervasive Computing, 2015.
- [8] Wang D L, Wang J G, Dong X B. *Research in Astronomy and Astrophysics*, 2012, 12(1): 509.
- [9] Subekti A, Rachmana N, Suksmono A B. *Physical Review B*, 2014, 12(3): 3895.
- [10] Hanabata Y, Katagiri H, Hewitt J W. *Astrophysical Journal*, 2014, 786(2): 86.

A Processing Method for Low SNR Repetitive Observation Spectrum

LIU Yuan-yuan¹, CHEN Jian-jun², QIU Bo^{1*}, FAN Xiao-dong¹, WEI Shi-ya¹, SONG Tao¹, DUAN Fu-qing^{3*}

1. Hebei University of Technology, Tianjin 300401, China

2. National Astronomical Observatories of Chinese Academy of Sciences, Beijing 100012, China

3. Beijing Normal University, Beijing 100875, China

Abstract At present, more than 7.6 million celestial spectra has been obtained by the LAMOST spectrum sky surveys, and the processing of low SNR spectra has been regarded as a difficulty in this domain. In this paper, a new method was proposed to deal with repeated observation spectra. The process of the method was: for every group of repetitive observation spectra, selecting that whose difference of red shift was within a certain range, and then an optimal stack algorithm based on SNR weighting was run to increase the SNR. The results of the processing of all repetitive observation spectra in LAMOST DR4 showed that this method was very effective for increasing the SNR for low SNR repetitive observation spectra. This method made the SNR of the 7 571 sets of stellar spectra reach the standard of parameter measurement; The SNR of the 3 357 quasar and Galaxy spectra improved at an average improvement rate of 56.38%; And 43 021 binary candidate candidates were obtained.

Keywords Repetitive observation spectrum; Low SNR; CCD noise sources; Candidate for binary star

(Received Aug. 14, 2017; accepted Dec. 29, 2017)

* Corresponding authors