

文章编号: 0253-2239(2008)06-1101-05

基于谱线检测的发射线星自动识别方法

刘中田¹ 邱宽民¹ 杨金福²

(¹ 北京交通大学电子信息工程学院自动控制工程系, 北京 100044)
(² 北京工业大学电子信息与控制工程学院自动化系, 北京 100022)

摘要 针对我国即将建成的大型巡天项目(LAMOST),给出了一种基于谱线检测的发射线星自动识别方法。主要步骤:1)通过获取谱线特征匹配值进行恒星谱线整体估计;2)利用提取出的恒星谱线特征检测恒星的巴耳末(Balmer)线;3)对获取的特征匹配值采用阈值法,并结合恒星 Balmer 线的检测结果,进行发射线星判别。通过对 SDSS DR4 所有光谱进行识别,共获得了 242 条具有恒星发射线的特殊天体。根据星表查询结果,这些天体包括发射线星、激变变星和一些未知特殊天体等。大量真实光谱数据实验表明,本文方法可有效识别发射线星。

关键词 模式识别;特殊天体;发射线星;谱线检测;小波变换

中图分类号 TP391.4 文献标识码 A

A Method for Automatic Recognition of Emission-Line Stars Based on Detection of Spectral Lines

Liu Zhongtian¹ Qiu Kuanmin¹ Yang Jinfu²

¹Department of Automatic Control Engineering, School of Electronics and Information Engineering,
Beijing Jiaotong University, Beijing 100044, China
²Department of Automatic Control, College of Electronic Information and Control Engineering,
Beijing University of Technology, Beijing 100022, China

Abstract Aiming at the LAMOST project, the largest sky survey project implemented in China, a novel method for automatic recognition of emission-line stars (ELS) based on detection of spectral lines is proposed. This method consists of three main steps: 1) the integral information of spectral lines is calculated by obtaining the correlations of the features of spectral lines; 2) the stellar Balmer lines are detected based on the features of stellar spectral lines; 3) the ELSs are automatically recognized by using the threshold of the correlations and the results of the stellar Balmer lines. In the experiments with all observed spectra from the sloan digital sky survey (SDSS) DR4, 242 peculiar celestial bodies with the stellar emission lines are obtained. According to the Simbad reference database, they are the celestial bodies include ELSs, cataclysmic variable stars (CVS), and unknown peculiar celestial bodies, etc. . The extensive experiments with observed spectra show that the proposed method can automatically recognize ELS effectively.

Key words pattern recognition; peculiar celestial bodies; emission-line stars (ELS); detection of spectral lines; wavelet transform

1 引言

恒星光谱中一般含有恒星吸收线或者吸收带^[1,2],只有特殊的恒星,其光谱中才会含有较强的发射线。本文主要讨论银河系内和近邻星系中的恒星^[3],它们的视向速度相对较小,红移值可认为是零。恒星的谱线波长位置相对固定,如果这些位置

上存在较强的发射线,则该天体很可能是发射线星(ELS)。发射线星相对来说非常稀少,它们在恒星演化过程中处于特殊的时期,因此天文学家对此类天体比较感兴趣。本文针对我国即将建成的大天区面积多目标光纤光谱望远镜(LAMOST)项目,旨在探索研究发射线星自动识别的有效方法。

收稿日期: 2007-08-07; 收到修改稿日期: 2007-12-28

基金项目: 国家 863 计划(2003AA133060)和北京交通大学科技基金(2006RC028)资助课题。

作者简介: 刘中田(1979-),男,讲师,主要从事模式识别、信号处理等方面的研究。E-mail: liuzht@bjtu.edu.cn

覃冬梅等^[4]应用主分量分析法结合最近邻分类方法,对恒星光谱型和光谱次型进行自动识别。通过设置合适的最近邻阈值,从恒星光谱中获得一些离群数据。实验发现,这些离群数据主要是光谱信噪比较低造成的误判,该方法并不能有效地识别发射线星。目前,尚未见到关于对发射线星有效的自动识别方法的报道。由于发射线星数量非常少,光谱数据空间离散较大,很难利用基于空间划分的传统模式分类方法进行有效识别。在海量光谱库中自动识别数量稀少的一类天体,需要综合考虑正确识别率和对其他天体的误识别率,也就是说,在保证正确识别率较高的前提下,尽量降低对其他天体的误识别率,以防止统计数据量相差悬殊造成的对自动识别的不利影响。

本文在充分研究发射线星特征的基础上,提出了一种基于谱线检测的发射线星自动识别方法。通过获取谱线特征匹配值^[5]进行恒星谱线整体估计,以克服低信噪比造成的不利影响。为了降低对其他天体的误识别率,进行恒星巴耳末(Balmer)线检测,以综合判别发射线星。该方法的优点在于:1)谱线特征提取方法^[5]抗噪声能力较强,获取的特征匹配值能够初步估计出谱线整体的发射与吸收性;2)在保证发射线星较高识别率的情况下,对其他天体的误识别率较低,低于1%。通过对SDSS DR4所有光谱进行实验,共获得了242条具有恒星发射线的特殊天体,其中,星表中未知的特殊天体具有重要的研究价值。大量真实光谱数据实验表明,本文方法可有效识别发射线星,并符合LAMOST光谱数据的要求。

2 恒星谱线整体检测方法

本方法基于谱线检测,目的是自动识别发射线星,所用到谱线检测方法主要有整体谱线检测和固定位置处谱线检测。整体谱线检测方法,主要包括谱线特征提取方法、特征模板构造方法以及相关匹配值计算方法^[5]。对已有方法进行继承性应用,有利于构成综合的恒星光谱自动识别系统,不仅能够识别正常恒星,而且能够识别特殊光谱恒星。

设 $f(n)$ 是原始光谱信号,按如下步骤对已有方法进行继承性应用:

1) 利用谱线特征提取方法得到 $f'_i(n)$, $i=2,3$,对应第2,3层小波特征^[6~8]。该谱线特征提取方法同时适用于吸收线和发射线的特征提取。

2) 获取与谱线特征模板相关匹配值中的最大

值 η^m 和最小值 η^n 。这里,选择与已有方法相同的谱线特征模板,主要有两个原因:a)发射线星光谱较少,构造模板比较困难;b)基于谱线特征匹配,利用相同模板既能够很好识别正常恒星,又能够有效识别发射线星。

3) 利用获得的 η^m 和 η^n 进行分析。谱线特征模板是由正常恒星光谱数据经过谱线特征提取后构成的。如果待测光谱属于发射线星,由于发射线的小波特征和吸收线的小波特征近似互为相反数^[5],则 η^m 和 η^n 一般满足如下关系:

$$|\eta^m| < |\eta^n|, \quad (1)$$

此时 $\eta^n < 0$,说明光谱中的发射线占主要成分,并且 $|\eta^n|$ 值越大,发射线强度越强。为了克服噪声对谱线检测的影响,通常需要设置一个阈值 σ ,构成如下判别式:

$$|\eta^m| < |\eta^n|, \quad \text{且} \quad |\eta^n| > \sigma. \quad (2)$$

结合固定位置的恒星谱线检测,就能够有效地识别发射线星。

3 固定位置谱线检测方法

利用谱线特征提取方法提取谱线的小波特征,恒星谱线的第2,3层小波特征如图1所示。

由图1可以看出,经过谱线的小波特征提取过后,噪声影响大大减小,在固定位置处检测谱线变得比较容易。

用如下算法检测固定波长值的谱线,假设谱线中心位置波长值为 λ 。

1) 令原始恒星光谱为 $f(n)$,经过谱线的小波特征提取后,第二层特征信号为 $f'_2(n)$,第三层特征信号为 $f'_3(n)$;

2) 利用信号波长对应关系,计算波长值 λ 对应的特征信号位置

$$n_\lambda = \text{round}[(\lambda - \lambda_s) / \lambda_c + 1], \quad (3)$$

式中 λ_s 是光谱信号的起始波长值, λ_c 是光谱信号的分辨率,round表示四舍五入取整运算;

3) 令 $i=1$ 和 $j=0$,这里 i 和 j 都是用于计数的整形变量,目的是可以更好地控制算法流程;

4) 如果 $i=1$,令 $t(n)=f'_3(n)$ 且阈值 $\text{signal}=3\text{sqrt}(1/n)$;如果 $i=2$,令 $t(n)=f'_2(n)$ 且 $\text{signal}=2\text{sqrt}(1/n)$;这里 n 为数据点数,sqrt为求取均方差运算;

5) 先获取特征信号 $t(n)$ 的 n_λ 附近区间的信号 $g(n_\lambda)$,区间选择 $[n_\lambda - \sigma', n_\lambda + \sigma']$, σ' 是阈值,通常选择4到10的值,依据谱线宽度和光谱分辨率而定,

取值较宽松；

6) 求取信号 $|g(n_x)|$ 的最大值 Z_m 和最大值的位
置 Z_w , 判断最大值 Z_m 是否大于阈值 signal , 最
大值的位置 Z_w 是否大于 $\sigma'-2$ 且小于 $\sigma'+2$, 若这些
条件满足, 令 $j=1$, 且进入第 7 步; 否则, 如果 $i=1$,

令 $i=2$, 且进入第 4 步;

7) 如果 $j=1$, 判断信号 $g(n_x)$ 在位置 Z_w 处的
值的正负号, 若为正值, 则判别波长 λ 处有发射线,
若为负值, 则判别波长 λ 处有吸收线, 计算中止; 否
则, 如果 $j=0$, 判别波长 λ 处没有谱线, 计算中止。

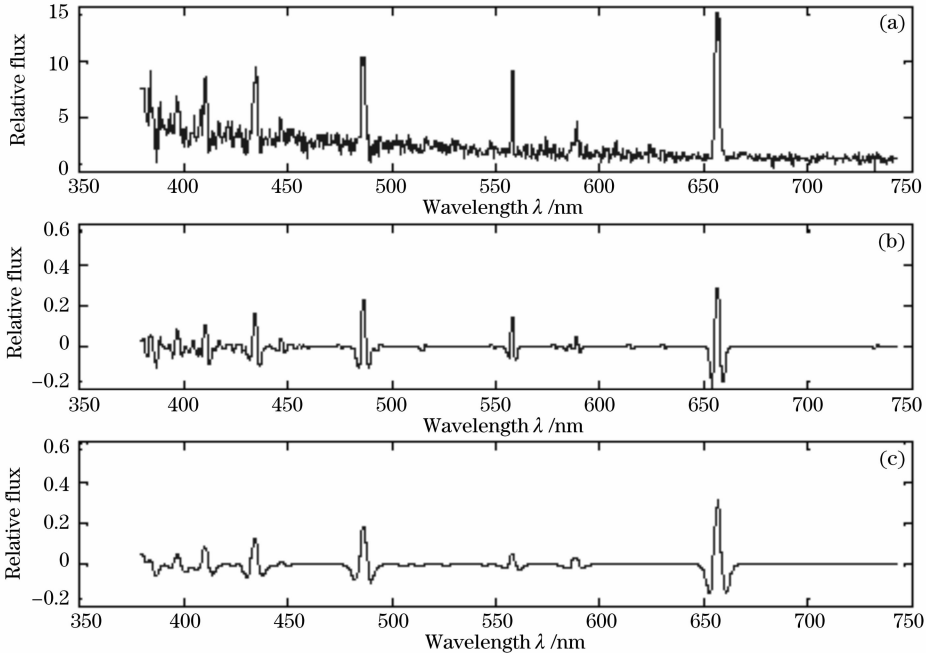


图 1 原始光谱(a), 第 2 层小波特征(b), 第 3 层小波特征(c)

Fig.1 Original spectrum (a), wavelet feature of the second scale (b), wavelet feature of the third scale (c)

对发射线星光谱的研究发现, 绝大多数发射线
星光谱主要存在 Balmer 线系的发射线。因此, 本文
所检测的固定位置谱线主要为 Balmer 线系的 H_α ,
 H_β 和 H_γ 线, 其波长分别为 656.3 nm, 486 nm 和
434 nm。其中, H_α 线较强, 受噪声影响较小, 一般
能够正确判断; H_β 和 H_γ 线谱线相对较弱, 噪声影
响较大时, 这两条谱线很可能会误判; 所以, 为了尽
可能提高发射线星识别率, 减小对其他天体的误识
别率, 针对 Balmer 线检测, 采用如下的判别规则。
如果 H_α 线为发射线, 且 H_β 和 H_γ 线至少有一条为
发射线, 则判别该天体为发射线星, 否则判别为其他
类型天体。

4 发射线星的识别

利用谱线整体检测所得的判别式结合 Balmer
线检测所制定的判别规则能够有效地识别发射线
星, 具体的识别流程如图 2 所示。

虚线以上的步骤在正常恒星识别方法中介绍
过^[5], 这里主要介绍虚线以下的部分。利用 ηm 、 ηn
和阈值 σ 构成第一个判断节点, 按照如下步骤进行

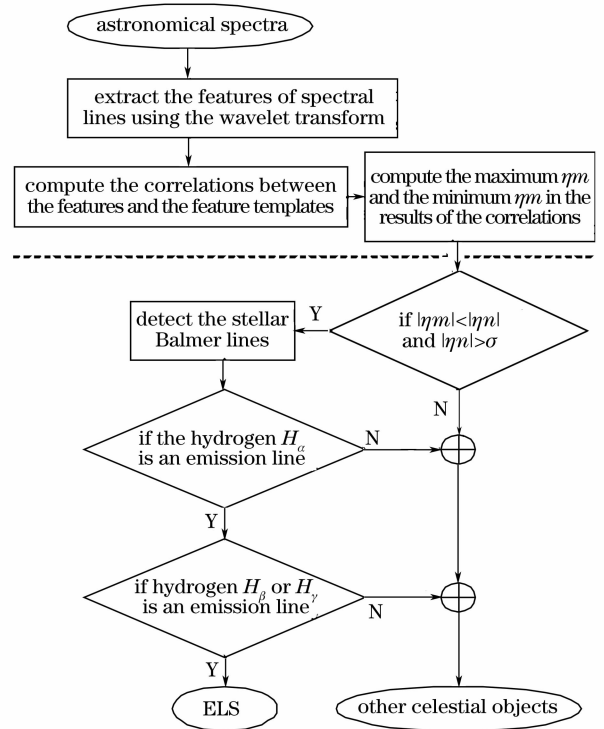


图 2 发射线星识别流程

Fig.2 Flow chart of recognizing emission-line stars (ELS)

发射线星判定:

1) 如果 $|\eta_m| < |\eta_n|$ 且 $|\eta_n| > \sigma$, 进行恒星氢线检测; 否则, 判别为其他类型天体;

2) 如果 $H\alpha$ 线是发射线, 则进入下一步; 如果 $H\alpha$ 线不是发射线, 则判别为其他类型天体;

3) 检测 $H\beta$ 和 $H\gamma$ 线, 如果有一条线是发射线, 则判别该天体为发射线星, 否则判别该天体为其他类型天体。

以上步骤可以看出, 该方法仍可能漏识别某些发射线星, 如只存在一条发射线的恒星光谱, 或由于某些原因, 导致 Balmer 线系发射线相对较弱, 不足以识别出。针对这些情况, 对该方法进行适当改进, 以适应某一种特殊天体的特征。例如, 可以只检测 $H\alpha$ 线是不是发射线来进行发射线星的判别, 这样

可以更大限度地识别出发射线星, 但是识别算法对其他天体的误识别率也会相应增加, 后续证认的工作量就会变得很大。

5 实验结果

对 SDSS DR4 中的所有光谱进行识别, 共获得 242 条具有氢发射线的特殊天体。实验发现, 阈值 σ 设置较宽松, 在 (0, 0.7) 范围内不断增加时, 对发射线星的正确识别率下降较小, 对其他类型天体的误识别率减小相对较快; 设置 $\sigma = 0.5$, 此时, 98% 以上的发射线星都能够被正确识别, 对其他类型天体的误识别率不大于 1%。图 3 为识别出的有代表性的四条特殊光谱。

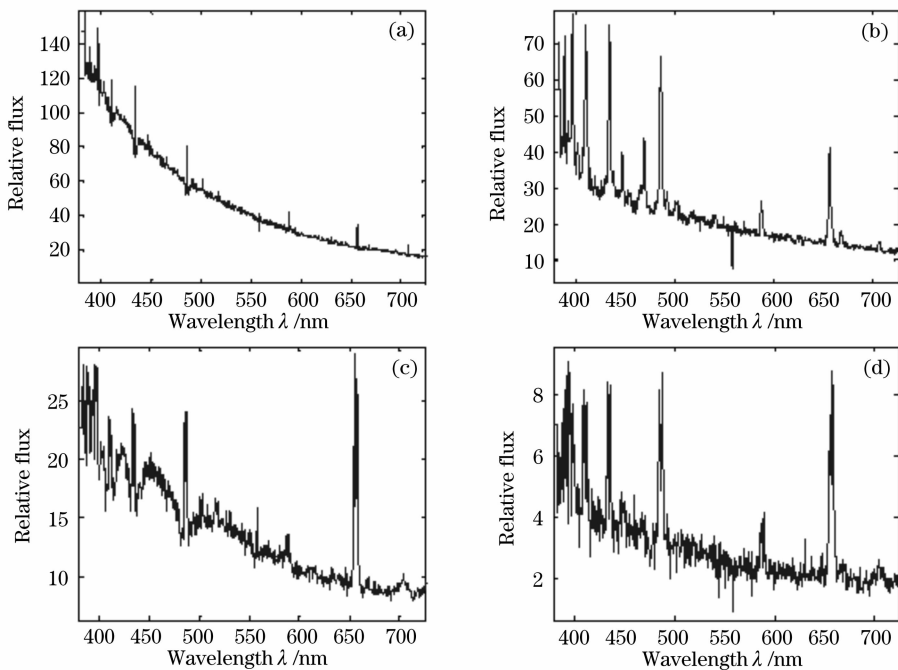


图 3 获得的部分特殊天体的光谱

Fig. 3 Spectra of the obtained peculiar celestial bodies

通过对 SDSS DR4 所有光谱进行实验, 共获得 242 条特殊天体^[5]。根据在 Simbad 星表中查询^[9]的结果, 这些天体包括发射线星、激变变星和一些未知特殊天体等。天文学家对未知特殊天体非常感兴趣, 例如, Liang 等^[10]根据本文识别结果给出了一个贫金属的近邻星系的证认结果。表 1 列出了部分识别出的天体。其中, 光谱文件位置用于在 SDSS DR4 光谱库中找出该天体, 由天区号和文件序号组合而成; 赤经 (R. A.) 和赤维 (DEC) 在 SDSS 光谱文件中已给出, 用于在 Simbad 星表数据库中进行查询。

6 结 论

提出了一种基于谱线检测的发射线星自动识别方法, 该方法利用恒星谱线整体估计和恒星 Balmer 线的检测结果综合判别发射线星。通过对 SDSS DR4 所有光谱进行识别, 获得了 242 条具有恒星发射线的特殊天体, 其中, 星表中未知的特殊天体具有重要的研究价值。大量真实光谱数据实验表明, 该方法可有效识别发射线星, 并符合 LAMOST 光谱数据的要求。

表 1 获得的部分特殊天体

Table 1 Obtained peculiar celestial bodies

Spectral location in SDSS DR4	R. A.	DEC
0306-004	218.7509	-7.684390
0343-198	238.8833	-1.819439
0346-263	243.3857	5.862282
0348-103	249.3426	-3.325358
0465-566	62.55695	-4.918056
0475-066	143.1592	1.150695
0495-238	192.5994	66.92374
0502-444	152.6544	2.820827
0543-065	118.1685	36.47312
0768-323	143.6035	53.81530
0770-358	147.8967	60.49434
0776-381	173.0646	62.81679
0781-093	191.1094	61.58740
0792-044	225.4051	55.02317
0831-410	136.2170	44.04873
0844-423	184.0293	5.337203
0846-595	190.9905	5.213305
0888-378	114.5739	28.92213
0958-297	196.3114	58.4823
0977-402	257.9379	30.22222
1205-188	120.7663	25.27415
1301-278	137.864	8.694665
1395-045	217.4828	41.75467
1592-207	139.2912	31.71878
1758-213	126.2381	7.617337

参 考 文 献

- Liu Zhongtian, Zhao Ruizhen, Zhao Yongheng *et al.*. A wavelet transform based method for the automatic detection of late-type stars[J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(7): 1158~1161
刘中田, 赵瑞珍, 赵永恒 等. 一种基于小波变换的晚型星自动识别方法[J]. *光谱学与光谱分析*, 2005, **25**(7): 1158~1161
- Liu Zhongtian, Li Xiangru, Wu Fuchao *et al.*. A method for auto-recognizing the M-type stars based on wavelet feature[J]. *Acta Electronica Sinica*, 2007, **35**(1): 157~160
刘中田, 李乡儒, 吴福朝 等. 基于小波特征的 M 型星自动识别方法[J]. *电子学报*, 2007, **35**(1): 157~160
- Carlos Jaschek, Mercedes Jaschek. *The Classification of Stars [M]*. Cambridge: Cambridge University Press, 1987
- Qin Dongmei, Hu Zhanyi, Zhao Yongheng. A PCA based efficient stellar spectra classification method[J]. *Spectroscopy and Spectral Analysis*, 2003, **23**(1): 182~186
覃冬梅, 胡占义, 赵永恒. 一种基于主分量分析的恒星光谱快速分类[J]. *光谱学与光谱分析*, 2003, **23**(1): 182~186
- Liu Zhongtian. *Automatic Recognition and Classification of Stellar Spectra [D]*. Beijing: Institute of Automation, Chinese Academy of Sciences, 2006
刘中田. *恒星光谱的自动识别与分类方法研究 [D]*. 北京: 中国科学院自动化研究所, 2006
- David L. Donoho. De-noising by soft-thresholding[J]. *IEEE Trans on Information Theory*, 1995, **41**(3): 613~627
- Zhou Xian, Liu Yilun. A study of automatic extraction method of defects on X-image[J]. *Acta Optica Sinica*, 2006, **26**(7): 1016~1020
周 贤, 刘义伦. X 光图像中缺陷的自动提取方法研究[J]. *光学学报*, 2006, **26**(7): 1016~1020
- Li Suwen, Xie Pinhua, Li Yujin *et al.*. Wavelet transform based differential optical absorption spectroscopy data processing[J]. *Acta Optica Sinica*, 2006, **26**(11): 1601~1604
李素文, 谢品华, 李玉金 等. 基于小波变换的差分吸收光谱数据处理方法[J]. *光学学报*, 2006, **26**(11): 1601~1604
- Simbad reference database, <http://simbad.u-strasbg.fr/sim-fid.pl>, 2006
- Y. C. Liang, J. Y. Hu, F. S. Liu *et al.*. SDSS J121811.0+465501.2: A new low surface brightness galaxy with low metallicity[J]. *The Astronomical J.*, 2007, **134**(8): 759~765