

基于 SDSS 的河内恒星 DIBs 自动识别方法*

章博, 潘楚楚, 叶宸, 唐轶峻

(浙江工业大学理学院应用物理系, 浙江 杭州 310023)

摘要: 星际弥散带(Diffuse Interstellar Bands, DIBs)自发现以来已经经历了近百年的研究,但是至今仍然是天体光谱学上的一个未解之谜。针对 SDSS DR7 的光谱数据提出了一种星际弥散带特征自动识别方法。该方法基于谱线特征匹配,通过光谱流量限制的方法进行星际弥散带特征的自动识别。利用它可对相对定标的巡天光谱进行广泛的星际弥散带候选天体搜索,在海量光谱数据中获取更多具有星际弥散带特征的河内恒星。通过对 SDSS DR7 中位置相对合适的超过 300 个盘的天体光谱的遍历,已经得到了一系列具有星际弥散带特征的候选河内恒星,并且证明了该方法简单有效且具有鲁棒性。这为载体证认等工作提供了大量辅助数据,极大地推进了星际弥散带的研究。

关键词: 星际弥散带(DIBs); 恒星光谱识别; 谱线特征匹配; 流量限制; SDSS

中图分类号: P141.5 **文献标识码:** A **文章编号:** 1672-7673(2013)03-0301-07

星际弥散带(Diffuse Interstellar Bands, DIBs)起源于星际介质的宽吸收带,其发现最早追溯到 1922 年^[1]。目前人们已经探测到数百条星际弥散带,但是还没有一条 DIB 的载体(carrier)被完全证认^[2]。经过近百年的研究,星际弥散带的起源至今仍然是一个亟待解决的天体光谱学之谜。早期研究由于受到观测技术等条件限制,多数都集中于对亮星(星等 <9)和河内恒星的星际弥散带确认。近些年随着观测技术的发展和设备的完善,星际弥散带的探测目标已经从河内恒星延伸到河外暗弱天体,研究重心逐渐倾向于河外天体星际弥散带谱线的发现和高碳有机分子等可能载体的证认^[2]。而 SDSS 等世界上先进的巡天计划已经累积了海量的光谱数据,SDSS(DR7)为 1 640 960 个天体提供了光谱,其中包括 464 261 个恒星光谱^[3]。这为广泛地开展星际弥散带识别和数据挖掘工作提供了重要的线索和条件,如果能有一种针对海量光谱数据进行星际弥散带自动识别的方法,势必能把星际弥散带的光谱目标扩展到更广的范围,为载体证认等工作提供辅助数据,推进星际弥散带的各方面的研究。

在星际弥散带研究方面,文[2]在星际弥散带与 217.5 nm 消光驼峰的关系方面取得了卓有成效的结果。在光谱分析方面,文[4]分析了 42 个类星体的 SPITZER IRS 中红外光谱并测量了连续谱和各发射线强度;文[5]则基于数据仓库进行了星系光谱分类,得到了极好的测试结果。而在数据自动处理上,文[6]针对中小规模数量的 FITS 文件设计了管理器,实现对 FITS 文件的快速查找、预览以及相关处理,说明了光谱文件批处理的可行性;同时,文[7]在 LAMOST 天体光谱自动处理方面取得一系列研究成果,对基于多层小波变换谱的发射线星的自动识别方法以及基于光谱特征匹配的恒星识别方法做了深入的研究。这些工作对于本文工作的开展提供了有益的理论指导。然而,由于许多星际弥散带具有不对称的复杂轮廓^[8],因此需要一种简单直接的方法进行星际弥散带识别。本文提出一种有效的、基于光谱匹配的河内恒星星际弥散带特征自动识别方法。该方法利用对固定波长位置的归一化流量限制,能够从天体光谱中将具有星际弥散带特征的河内恒星谱线筛选出来。其优点在于:(1)简单有效,并对符合条件的目标结果返回星际弥散带中心波长值,后续能将结果按照类型分类和可视化,可行性强。(2)对于异常的数据,程序具有较好的自适应能力,具有鲁棒性。

* 基金项目:浙江省大学生科技创新活动计划(新苗人才计划)资助。

收稿日期:2012-06-20;修定日期:2012-07-04

作者简介:章博,男,本科。研究方向:天文技术与方法。Email: chamberlain31@163.com

1 星际弥散带自动识别原理

星际弥散带的确认主要基于星际弥散带所具有的两个基本属性^[9]:

(1) 多次独立观测中, 谱线中的吸收带具有恒定的日心视向速度;

(2) 吸收带的表观强度和轮廓必须具有较好的重复性, 即多次观测都能找到表观强度和轮廓特征一致的吸收带。

据此, 寻找具有星际弥散带特征的天体可以有两种方法:

(1) 对多次独立观测的光谱数据进行检测。光谱数据中均检测到吸收带且其日心不变视向速度恒定的天体可以确认为含有星际弥散带特征的天体。

(2) 基于固定位置的吸收带检测。使用已经确认的星际弥散带中强吸收带的中心波长作为固定位置, 对其近邻像素点进行流量检测, 若检测到存在吸收带, 则筛选出来作为星际弥散带候选天体, 待进一步确认。

图 1 所示的是几个具有 769.9 nm DIB 的恒星谱线。考虑到星际弥散带的谱线可能具有复杂的轮廓, 因此需要一种简单可行的检测方法。由于 SDSS 中的光谱数据中只有部分恒星的光谱具有重复观测数据, 故本方法主要基于第二种途径, 后期再对于结果进行重复观测数据的搜索和对比。自动识别软件对于检测的 fit 文件将给出赤经、赤纬、星等、红移等信息以及包含的可能为星际弥散带吸收带的位置中心波长。SDSS 提供的光谱数据像素点的波长间隔大约为 0.102 nm ~ 0.167 nm 之间, 因此吸收带检测的半峰全宽极限被限制在相近的数值, 即: 半峰全宽接近或者小于 0.1 nm 的星际弥散带将不可能被检测出来^[9]。所以实验中选取了半峰全宽较大、强度较强^[11]的几个星际弥散带作为检测目标, 具体中心波长为以下几个: 442.8 nm、570.5 nm、578.0 nm、579.7 nm、619.6 nm、620.3 nm、626.9 nm、628.3 nm、637.9 nm、661.3 nm、666.0 nm、699.3 nm、772.4 nm。

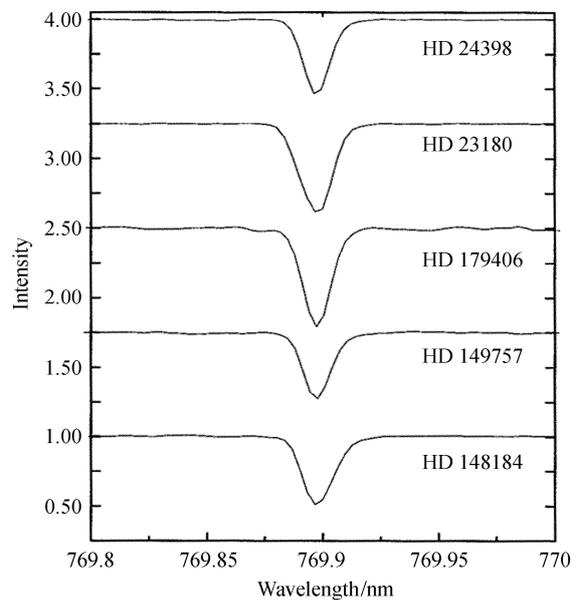


图 1 具有 769.9 nm DIB 的恒星谱线^[10]

Fig. 1 Spectra of several stars with the 769.9nm DIB feature

2 星际弥散带自动识别算法、流程

针对 SDSS 光谱数据中的强线选择、光谱数据信噪比、波长分辨率、FWHM 控制等方面设计识别过程的关键算法, 其主要目的是判断天体光谱是否具有星际弥散带特征, 并据此对光谱数据进行识别分类: 具有星际弥散带特征(返回中心波长值), 不具有星际弥散带特征(返回空)。

光谱的星际弥散带特征自动检测流程如图 2, 具体由以下步骤组成:

(1) 读取 fit 文件所包含的光谱数据(包括赤经、赤纬、星等、红移值、天体类型及归一化的光谱数据等);

(2) 根据天体类型, 筛选出恒星类型的天体光谱;

(3) 判断是否满足红移 $Z < 0.001$, 筛选出河内恒星, 作为星际弥散带检测对象;

(4) 对需要检测的几个星际弥散带波长进行固定位置吸收带检测。在该中心波长邻近的像素点取得归一化流量值, 检测要求该处最近邻像素点的归一化流量值低于 0.8, 次近邻像素点的归一化流量值低于 0.9;

(5) 按照检测的固定位置中心波长从小到大, 对每个 fit 光谱遍历所有需要检测的星际弥散带中心波长, 返回检测到的存在吸收带的中心波长值;

(6) 每一次检测对一个盘中所有的 fit 文件遍历, 并保存结果。

实验中选取的流量限制条件的数值 0.9 是一个经验参数, 根据文[9]中确认的星际弥散带的信息可以知道, 符合最近邻小于 0.8 和次近邻小于 0.9 的星际弥散带必为相当强的星际弥散带, 其检测结果的可靠程度是很高的, 这在一定程度上避免了低分辨率对星际弥散带识别造成的影响。

3 实验及分析

3.1 实验的理论预期

1. 光谱类型分布。调查了国外几项较大的星际弥散带研究^[12-15], 总共涉及了 510 个光谱目标, 并对其中在 SIMBAD 数据库中天体类型为 Star 并且光谱类型明确的恒星进行了统计, 得到了 288 颗星际弥散带特征恒星的光谱类型分布图(如图 3)。可以看出, 这些星际弥散带研究主要集中在 O、B、A 型恒星, 即早型星。

2. 候选天体光谱的可信度。考虑到星际弥散带的确认必须遵循两个原则, 并且由于星际弥散带较多地被发现存在于强红化星和早型星的光谱中^[1], 因此在实验中, 必须优先关注以下几类可信度较高的结果:

- (1) 有多次观测数据, 且各次观测的光谱中均检测到吸收带;
- (2) 单次观测数据, 但是检测到了吸收带, 且谱线较好;
- (3) 单次观测数据, 但是检测到了吸收带, 且属于早型星。

3.2 实验结果及其分析

实验中对 SDSS DR7 中的天区位置相对合适的 300 多个盘进行遍历, 共检测了超过 20 万个天体光谱, 得到了一系列具有星际弥散带特征的候选河内恒星, 并对这些结果进行了自动分类。检测到的具有星际弥散带特征的恒星目标共有 117 个, 在附录中给出了盘和在对应盘中检测到的星际弥散带特征候选恒星的星际弥散带波长, 表格中数字“1”代表在对应盘中检测到了该波长 DIB, 空格则表示没有检测到该波长 DIB。其中, 有 6 个波长的星际弥散带在检测过程中未检测到候选恒星, 则未列出; 若干个未检测到有任何候选恒星的盘也已经略去不列。

图 4 所示的是检测结果中较为可靠的谱线以及值得关注的谱线, 并在表 1 中给出了天体的部分信息。其中, 54463-2856-025 只具有单次观测数据, 但是谱线在 620.3 nm 位置处显示出较好的轮廓, 值得进一步验证。而另外两条谱线为同一目标的光谱, 其 619.6 nm 处的吸收带在强度和轮廓上显示出较好的重复性, 极有可能是具有 620.3 nm DIB 的河内恒星。

这 117 个星际弥散带候选恒星光谱类型包括 A0、F2、F5、F9、G2、K1、K3、K5、K7、M0、M0V、M1、M2、M3、M4 等, 与理论预期中统计的具有星际弥散带特征的 288 个恒星光谱类型分布相比明显偏晚。其原因主要有以下两个:

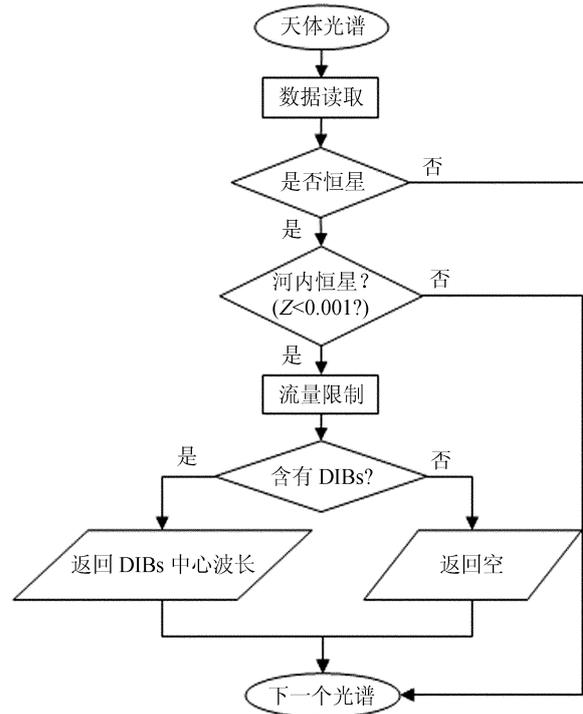


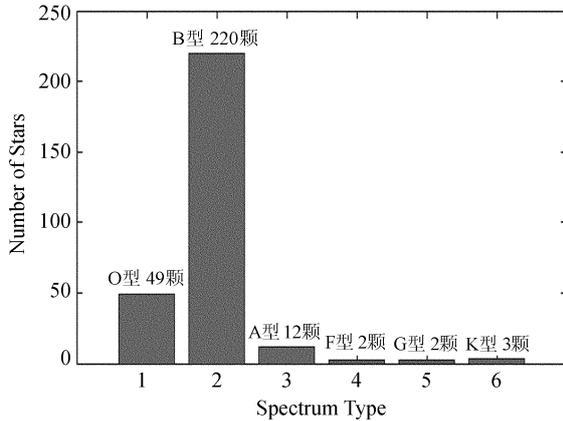
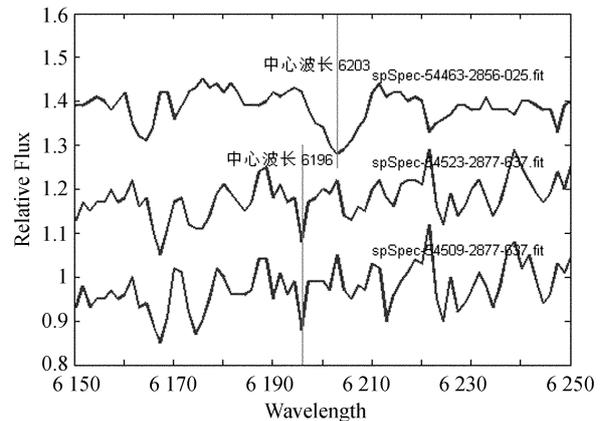
图 2 星际弥散带自动检测流程图

Fig. 2 Flowchart of our algorithm of selecting DIBs

表 1 图 4 中光谱对应的恒星信息

Table 1 Information of Stars Plotted in Fig. 4

plate	MJD	fiberid	RA	DEC	Z	MAG	DIBs	Spec Type
2 856	54 463	025	168.750 49	38.171 111	1.851 1E-5	16.472 481	6 203	F9
2 877	54 523	637	175.352 17	0.026 877	0.000 417 812	18.241 278	6 196	F9
2 877	54 509	637	175.352 17	0.026 877	0.000 431 909	18.240 673	6 196	F9

图 3 288 颗具有星际弥散带特征的恒星的光谱类型统计
Fig. 3 Statistics of spectral types of 288 stars with DIBs图 4 检测到的可信度较高的星际弥散带特征恒星光谱
Fig. 4 Spectra of stars with DIBs detected with high confidence levels

(1) 根据 Slyk^[10] 等人的分析, 要对弱星际弥散带轮廓的子结构 (substructure) 和微细结构 (fine structure) 进行可靠的研究, 所使用的光谱分辨率要达到 $R = 100\,000$ 。而 SDSS DR7 中的光谱分辨率 R 大约在 $1\,800 \sim 2\,200$, 与星际弥散带轮廓微细结构研究的要求相差很大, 使得用于判断吸收特征的像素点流量数据的个数减少, 得到的轮廓与高分辨率情况下的轮廓必然会有一些差别, 这导致了检测结果可靠性有一定的降低。针对这个问题, 选取 13 个强星际弥散带中心波长作为检测目标, 根据文[9]提到的波长分辨率与 DIB 识别能力的关系, 仍然可以保证在这 13 个可能出现较强星际弥散带 ($\text{FWHM} \gg 0.1 \text{ nm}$) 的波长位置的特征被有效识别出来。另外, 本文所做的工作本身与星际弥散带轮廓的微细结构研究亦有所不同, 对分辨率的要求也会有所降低, 并且通过控制光谱信噪比和利用重复观测数据提高了检测结果的可靠性。

(2) 根据对文[12-15]中涉及的 288 个光谱目标在天区中位置的统计(如图 5), 容易知道研究较多

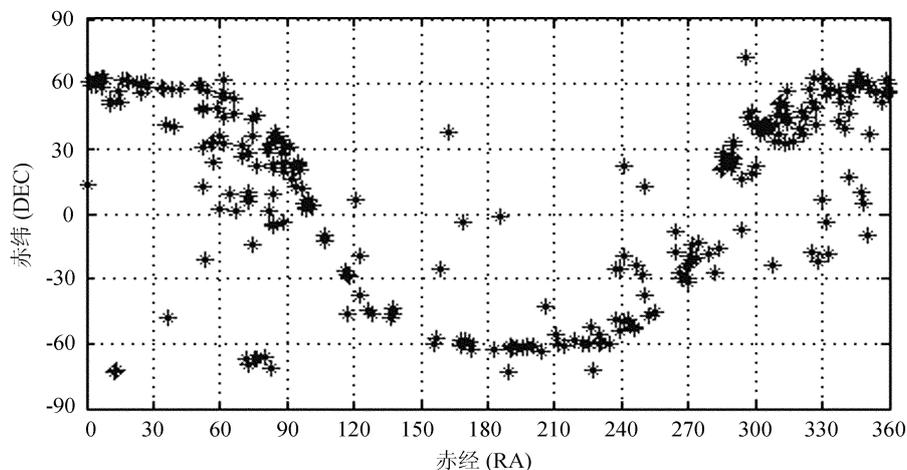


图 5 [12-14] 中的星际弥散带特征恒星的位置分布

Fig. 5 The surface distribution of stars with DIB features in Ref [12-14]

的星际弥散带是处于低银纬区域。但是 SDSS DR7 的主要观测天区分布在高银纬区域。银纬的高低与星际弥散带载体在星际空间的分布情况有直接的联系, 这是导致光谱类型分布与理论预期有差别的主要原因之一。

4 小 结

本文的创新之处在于将光谱特征匹配应用在对星际弥散带特征河内恒星的光谱数据挖掘, 为星际弥散带的载体证认等研究提供了大量辅助数据。通过实验, 证明了本文提出的自动识别河内恒星星际弥散带特征的可行性。对于接下来的工作方向, 一方面可以改进算法, 利用精度更高的处理方法去处理已有的光谱, 以得到更高的识别准确率。另外还可以利用高精度的标准恒星谱线进行星际弥散带的确认, 或者利用观测设备对只有单次观测数据的候选恒星进行观测验证等。另一方面, 可以增加星际弥散带的遍历样本, 如进行剩余 SDSS 光谱数据遍历, 寻找更多候选天体; 选择其他巡天计划的光谱数据和等待新数据的释放, 如 SDSS DR8 等, 提高被检测光谱的分辨率和信噪比, 增加自动识别星际弥散带候选恒星的可靠性。

致谢: 感谢国家天文台胡景耀研究员的有益指导和支持。

附录 检测结果列表
Appendix A list of results

plate	5705	5780	5797	6196	6203	6283	6379	plate	5705	5780	5797	6196	6203	6283	6379
277				1				2207	1						
279	1							2209		1					
280		1						2332		1			1		
323	1							2389	1						
338							1	2409					1		
467	1							2419	1						
507	1			1				2420	1						
525				1		1		2421	1						
565	1							2423	1						
574	1	1						2459	1				1		
844	1							2474	1						1
846						1		2528						1	
852			1					2532	1						
1041			1					2558	1						
1184	1							2559	1						
1284				1				2567	1						
1299		1						2569	1			1			
1324	1		1					2707						1	
1433	1				1	1		2807	1						
1568	1							2817	1						

续表

plate	5705	5780	5797	6196	6203	6283	6379	plate	5705	5780	5797	6196	6203	6283	6379
1569	1							2856						1	
1659	1	1						2861				1			
1667				1	1			2862				1			
1668							1	2871	1	1		1	1	1	1
1756	1							2876	1		1				1
1758	1						1	2877			1	1		1	
1849				1				2895		1		1			
2030							1	2897			1				
2056	1		1					2901				1			
2076		1			1			2913	1	1					
2077				1				2920						1	
2169				1				2922	1	1				1	
2183	1	1						2926	1			1		1	
2188	1	1	1				1	2967			1				
2202	1														

参考文献:

- [1] M L Heger. Further study of the sodium lines in class B stars; The Spectra of Certain Class B Stars in the Regions 5630A-6680A and 3280A-3380A; Note on the Spectrum of $[\gamma]$ Cassiopeiae between 5860A and 6600A. Lick Observatory Bull. 1922, 10, 337: 141-145.
- [2] 向福元, 梁顺林, 李爱根. 星际弥散吸收带 [J]. 中国科学 G 辑: 物理学 力学 天文学, 2009, 39(4): 481-493.
Xiang Fuyuan, Liang Shunlin, Li Aigen. Diffuse and absorb interstellar bands [J]. Science in China (Series G: Physics, Mechanics & Astronomy), 2009, 39(4): 481-493.
- [3] SDSS Data [DB/OL]. [2012-06-20]. <http://www.sdss.org/dr7/>.
- [4] 谢艳霞, 商朝晖. 类星体的中红外光谱分析 [J]. 天文研究与技术——国家天文台台刊, 2011, 8(3): 205-212.
Xie Yanxia, Shang Zhaohui. An analysis of quasar mid-infrared spectra [J]. Astronomical Research & Technology——Publications of National Astronomical Observatories of China, 2011, 8(3): 205-212.
- [5] 孙士卫, 罗阿理, 张继福. 基于数据仓库的星系光谱分类 [J]. 天文研究与技术——国家天文台台刊, 2007, 4(3): 276-282.
Sun Shiwei, Luo Ali, Zhang Jifu. Galaxy spectral classification Based on data warehouse [J]. Astronomical Research & Technology——Publications of National Astronomical Observatories of China, 2007, 4(3): 276-282.
- [6] 樊东卫, 崔辰州, 赵永恒. FITS 文件管理器设计与实现 [J]. 天文研究与技术——国家天文台台刊, 2011, 8(3): 306-314.

- Fan Dongwei, Cui Chenzhou, Zhao Yongheng. Design and implementation of a FITS file manager [J]. *Astronomical Research & Technology—Publications of National Astronomical Observatories of China*, 2011, 8(3): 306–314.
- [7] 刘中田, 邱宽民. 基于光谱特征的恒星自动识别方法 [J]. *光谱学与光谱分析*, 2010, 30(1): 274–277.
- Liu Zhongtian, Qiu Kuanmin. A method for auto-recognizing the stars based on spectral feature [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(1): 274–277.
- [8] T P Snow. The unidentified diffuse interstellar bands as evidence for large molecules in interstellar medium [J]. *Spectrochimica Acta Part A*, 2001, 57: 615–626.
- [9] L M Hobbs, D G Yorks, T P Snow, et al. A catalog of diffuse interstellar bands in the spectrum of HD 204827 [J]. *The Astrophysical Journal*, 2008, 680(3): 1256–1270.
- [10] K Skyk, G A Galazutdinov, F A Musaev, et al. A search for fine structure inside high resolution profiles of weak diffuse interstellar bands [J]. *Astronomy and Astrophysics*, 2006, 448(1): 221–229.
- [11] P Jenniskens, F X Desert. A survey of diffuse interstellar bands (3800–8680 Å) [J]. *Astronomy and Astrophysics Supplement Series*, 1994, 106(1): 39–78.
- [12] G A H Walker. Photoelectric measures of the 4430Å diffuse interstellar band [J]. *Monthly Notices of the Royal Astronomical Society*, 1963, 125(1): 141–167.
- [13] E J Wampler. Scanner observation of λ 4430 [J]. *The Astrophysical Journal*, 1966, 144(3): 921–937.
- [14] G H Herbig. The diffuse interstellar bands. IV. the region 4400–6850 Å [J]. *The Astrophysical Journal*, 1975, 196(1): 129–160.
- [15] G H Herbig. The diffuse interstellar bands. IX. constraints on the identification [J]. *The Astrophysical Journal*, 1993, 407(1): 142–156.

An Automated Method of Selecting Galactic Stars with DIB Features from the SDSS

Zhang Bo, Pan Chuchu, Ye Chen, Tang Yijun

(Department of Applied Physics, Zhejiang University of Technology, Hangzhou 310023, China, Email: chamberlain31@163.com)

Abstract: The problem of Diffuse Interstellar Bands (DIBs) has been studied for almost a century since the discovery, but it still remains a mystery. A method is proposed in this paper for automatically selecting Galactic stars with DIB features from the SDSS DR7 data. This method is based on spectral-feature matching, and selecting of DIB features by requiring sufficiently low relative neighboring-pixel fluxes for the central-wavelength pixels of DIB lines (—line-profile prominence). We are able to select candidates of Galactic stars with DIB features with this method from calibrated spectra, as shown by experiments on over 300 SDSS plates. The paper thus shows the simplicity, effectiveness, and robustness of the method. The method will provide a huge amount of data for research areas such as identifications of stars of these features, and will push forward research of DIBs.

Key words: Diffuse Interstellar Bands (DIBs); Spectra recognizing of stars; Feature matching of spectral lines; Relative flux confining; SDSS