

基于小波特征的 M 型星自动识别方法

刘中田¹, 李乡儒², 吴福朝², 赵永恒³

(1. 北京交通大学电子信息工程学院, 北京 100044;

2. 中国科学院自动化研究所模式识别国家重点实验室, 北京 100080; 3. 中国科学院国家天文台, 北京 100012)

摘 要: M 型星对研究恒星的演化具有重要价值. 在我国正在实施的大型巡天项目(LAMOST 项目)中, 急需 M 型星的自动识别系统. 本文给出了一种自动识别 M 型星的新方法, 该方法由以下主要步骤组成: ①选取一定波长范围的光谱进行 5 层小波变换, 从第 5 层小波系数中提取出小波特征; ②利用小波特征检测 M 型星特征频率和吸收带位置; ③根据特征频率和吸收带位置的检测结果进行 M 型星识别. 大量真实光谱数据实验表明, 本文方法十分有效, 识别率高达 97.56%.

关键词: 天体光谱; 吸收带检测; 特征频率; 小波特征; 小波变换

中图分类号: TN911.7 **文献标识码:** A **文章编号:** 0372-2112(2007)01-0157-04

A Method for Auto-Recognizing the M-type Stars Based on Wavelet Feature

LIU Zhong-tian¹, LI Xiang-ru², WU Fu-chao², ZHAO Yong-heng³

(1. School of Electronics and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China;

3. National Astronomical Observatory, Chinese Academy of Sciences, Beijing 100012, China)

Abstract: M-type stars play a significant role in the study of star evolution. The LAMOST project, the largest sky survey project being implemented in China, urgently needs a system for auto-recognizing M-type stars. This paper presents a novel method that can automatically recognize M-type stars. This method consists of three main steps: First, after a wavelet transform with 5 scales on the spectra in a selected wavelength region, the wavelet features are extracted from the transformed coefficients on the 5th scale. Then, the characteristic frequency of M-type stars and the locations of absorption bands are obtained accurately through the wavelet features. Finally, based on the results of the former step, M-type stars in all kinds of celestial bodies can be recognized. The extensive experiments with real observed spectra show that the proposed method is effective and the correct rate of this method is as high as 97.56%.

Key words: astronomical spectra; detection of absorption bands; characteristic frequency; wavelet feature; wavelet transform

1 引言

人们普遍认为, 恒星的连续谱主要是由恒星的表面有效温度、重力加速度和化学丰度决定的. M 型星的表面有效温度较低, 一般低于 3600K, 星的颜色呈现为红色, 其光谱特征主要表现在大量的分子吸收带上, 其中, 以 TiO 分子吸收带为主^[1]. M 型星是银河系中数量最多的一类恒星, 它位于恒星的哈佛系统分类^[2]序列的末端, 几乎所有的主序星都要演化到这一类型, 因此, 它对研究恒星的演化具有重要价值^[1]. 我国正在实施的 LAMOST 巡天计划, 在利用巡天成像数据选源时, 容易将类星体误认为是恒星而加以观测, 同时也会混入其他类型天体. 因此, 本文需要在巡天所得光谱中研究 M 型星的识别问题, 并要求所用方法能较好地地区分类星体.

目前对恒星子型的识别主要是从光谱型和光谱次型进行分析的. Kutz 在他的博士论文中^[2], 应用交叉相关认证和主分量分析方法, 将 MK 分类方法应用到低分辨率光谱(1.4 nm). 覃冬梅等应用径向神经网络的组合模型, 对恒星光谱次型进行自动识别^[3]. 这些方法都可以从恒星中识别出 M 型星, 但是不能用于巡天观测的各种类型天体中. Sloan Digital Sky Survey(SDSS)利用一系列模板, 通过交叉认证的方法识别过晚型星. 对于 LAMOST 项目来说, 由于流量不定标, 很难建立一套实用的模板, 基于模板交叉认证的方法不适用. 另外, 类星体是目前可观测到的红移最大、距离最远的天体, 其光谱的信噪比一般较低^[4]. M 型星光谱的信噪比一般也较低, 而且具有和类星体光谱类似的宽发射峰, 这就增加了正确识别 M 型星的困难.

本文通过检测强 TiO 分子吸收带来进行 M 型星的识别。为了克服信噪比较低的不利因素, 本文选取一定波长范围的光谱进行 5 层小波变换, 第 5 层小波系数包含 M 型星吸收带的主要特征。对第 5 层小波系数进行频谱分析, 可以得到 M 型星的特征频率, 然后结合小波特征进行吸收带位置检测, 根据检测结果设计决策规则, 从而将 M 型星从众多类型天体中识别出来。本文方法的优点在于: (1) 直接利用 M 型星分子吸收带的特征进行 M 型星识别, 不依赖模板匹配, 提高了识别算法的速度; (2) 选择合适的光谱波长范围和分子吸收带, 减小了噪声对 M 型星识别的影响, 识别正确率较高, 算法鲁棒性较强。方法的缺点主要在于, 对分子吸收带不明显的较早型 M 型星识别效果较差。

2 M 型星的小波特征

M 型星光谱特征主要表现在大量的分子吸收带上, 其中以 TiO 分子吸收带为主。在 TiO 分子吸收带中, $\lambda_{685.2}$ 、 $\lambda_{712.6}$ 和 $\lambda_{767.2}$ 处的吸收带较强, 几乎所有的 M 型星光谱都具有这些吸收带。由于 $\lambda_{686.7}$ 处有较强的大气吸收带, $\lambda_{685.2}$ 处的吸收带一般会与 $\lambda_{686.7}$ 处的吸收带混合形成较强的分子吸收带^[1]。本文中, 符号 λ 表示波长, 单位为纳米 (nm)。

蓝端和红外端光谱的信噪比一般较低, 这些范围内的吸收带较弱, 多数吸收带只在部分 M 型星光谱中存在, 而且信噪比低时不易检测, 所以本文只选择了 $\lambda_{623}-\lambda_{846}$ 范围的光谱进行处理。光谱分辨率为 0.5nm, 所得信号表示为 $f(n)$, $n=0, 1, \dots, 446$ 。

小波变换^[5]是信号处理的一种研究方法, 能够对信号进行多分辨分析。为了算法描述的方便, 先引进如下定义:

定义 1: 对信号进行 i 层小波变换得到第 i 层高通小波系数的算子用 $W_i(\cdot)$ 表示, 而算子 $\tilde{W}_i^{-1}(\cdot)$ 表示对第 i 层高通小波系数进行小波反变换。

定义 2: 算子 $\Delta(\cdot)$ 表示对信号进行前项差分。前项差分计算如下:

$$\bar{f}(i) = \Delta(f(i)) = f(i+1) - f(i) \quad (1)$$

其 $i=0, 1, \dots, 445$ 中。

首先对信号 $f(n)$ 进行 5 层小波变换^[6], 然后对第 5 层高频小波系数 $w_5(n)$ 进行小波反变换得到 $f_5(n)$, 最后对 $f_5(n)$ 做一次前项差分得到信号 $\bar{f}_5(n)$ 。变换过程如图 1 所示。

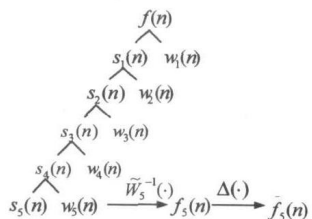


图 1 小波变换过程和前项差分

经过研究发现, 5 层小波变换可以较好地提取出 M 型星的吸收带特征, 4 层及以下小波变换不能很好地描述 M 型星的吸收带特征, 6 层及以上小波变换, 由于边缘效应的影响, 提取结果并不比 5 层变换更好, 有时可能更差。运算过程中采用了保持各层小波系数的点数不变的卷积型算法^[7]。

本文提取的小波特征可以表示为

$$w_5(n) = W_5(f(n)) \quad (2)$$

$$f_5(n) = \tilde{W}_5^{-1} W_5(f(n)) \quad (3)$$

其中 $n=0, 1, \dots, 446$ 。这里, 选取了两个变换结果作为特征, 主要是为了更好地检测吸收带。

3 吸收带检测及 M 型星识别

本文的吸收带检测主要分为两个部分: 1. 吸收带整体特征频率检测; 2. 吸收带具体位置检测。依据这两个部分的检测结果, 设计决策规则, 进行 M 型星识别。

3.1 特征频率检测

为了克服波长范围选择较大对特征频率检测的影响, 本文选择了光谱质量较好的 $\lambda_{623}-\lambda_{753}$ 范围, 也就是信号 $f(n)$ 的前 261 个值, 对应的小波特征为 $w_5(n)$, $n=0, 1, \dots, 260$ 。在所选波长范围内, 有两个强分子吸收带 $\lambda_{685.2}$ 和 $\lambda_{712.6}$, 小波特征 $w_5(n)$ 中相应的两个波谷最小值位置分别约为 λ_{664} 和 λ_{708} , 间隔约为 87, 相当于主要频率成分的周期 $T_z \approx 87$ 。

N 点离散傅里叶变换 (DFT) 可以表示为:

$$F(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) W_N^{nk} \quad (4)$$

式中, $k=0, 1, \dots, N-1$, $W_N = e^{-j2\pi/N}$ 称为蝶形因子^[8], $x(n)$ 是原信号, $N=261$ 。

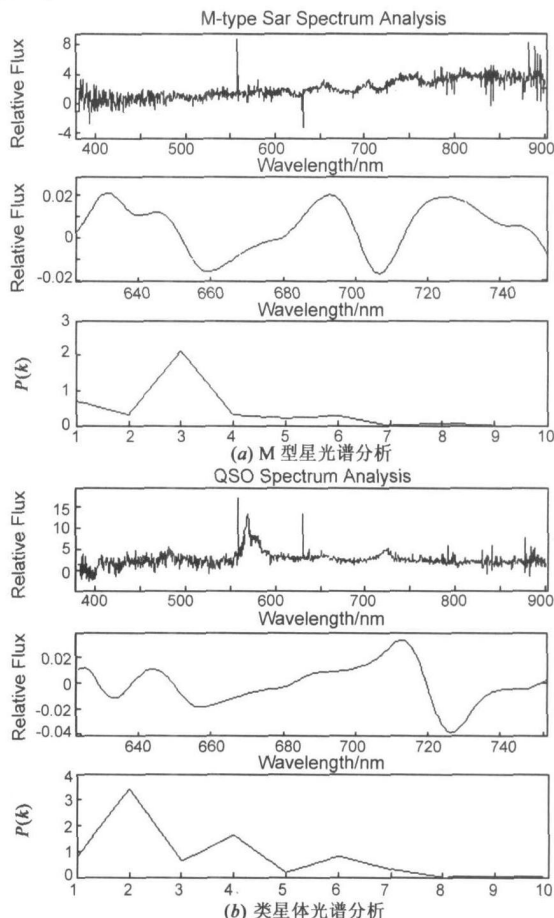


图 2 原始光谱 (上), 第 5 层小波系数 (中), 离散傅里叶变换后的频谱图 (下)

离散傅里叶变换后,第 i 个点对应的频率(去除了直流分量,频率最大值为 1)

$$F_i = \frac{i}{N}, N = 261, i = 1, 2, \dots, 130 \quad (5)$$

所以,主要频率成分的频率

$$f_z = \frac{1}{T_z} \approx \frac{3}{261} = \frac{3}{N}, N = 261 \quad (6)$$

即对应第 3 个点,本文称之为 M 型星的特征频率.图 2 表示的是 M 型星和类星体光谱的频率分析.

设 $P_i(k)$ 是第 i 条天体光谱 x_i 对应的频谱幅值,其中 $k = 1, 2, \dots, 130$, 去除了直流分量.特征频率的幅值占整个频段频谱幅值的百分比由下式计算:

$$\sigma(x_i) = \frac{P_i(3)}{\sum_{k=1}^{130} P_i(k)} \quad (7)$$

特征频率处有约束关系如下:

$$P_i(3) > P_i(2) + P_i(4) \quad (8)$$

$$\sigma(x_i) > \delta \quad (9)$$

其中, $\delta > 0$ 是阈值.为了阐述方便,本文称该约束关系为“约束规则 1”.

3.2 吸收带位置检测

本文需要检测的分子吸收带位置分别是 $\lambda_{85.2}$ 、 $\lambda_{712.6}$ 和 $\lambda_{767.2}$.由于这三个分子吸收带都比较强,在每个吸收带附近一般都形成了明显的波谷,这为在低分辨率和低信噪比的情况下检测吸收带提供了可能.

按公式(3)求取小波特征 $f_5(n)$,然后求取 $\bar{f}_5(n) = \Delta(f_5(n))$,本文是用信号 $-\bar{f}_5(n)$ 来检测 M 型星的吸收带位置的,具体的特征分析如图 3.

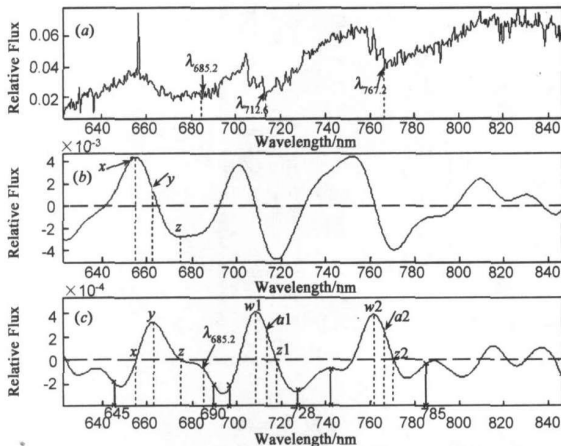


图 3 (a)M 型星原始光谱, (b)小波特征 $f_5(n)$, (c)信号 $-\bar{f}_5(n)$ 及吸收带检测

如图 3(a), 一般来说 $\lambda_{85.2}$ 附近的吸收带由于受大气吸收带的影响,吸收带引起的波谷比较宽,波谷最小值位置变化范围也比较大.图 3(b)和(c)中的 x, y, z 点分别是对应点,实验表明, x 点检测结果较理想,可以用该点来推测 $\lambda_{85.2}$ 处吸收带的位置.对于 $\lambda_{712.6}$ 和 $\lambda_{767.2}$ 处的吸收带,可以通过如下方法较准确地检测出来:如图 3(c)所示,首先在 $\lambda_{69.7} - \lambda_{728}$ 和 $\lambda_{742} - \lambda_{785}$ 范围内分别检测最大值点位置 w_1 和 w_2 , 然后向右

分别检测第一个零交叉点^[9]位置 z_1 和 z_2 , 则吸收带位置计算如下

$$a_i = (w_i + z_i)/2, i = 1, 2 \quad (10)$$

其中, $i = 1, 2$ 分别对应吸收带位置 $\lambda_{712.6}$ 和 $\lambda_{767.2}$.

对 SDSS DR2 中的 5408 条 M 型星光谱用本文方法进行吸收带位置检测,结果如图 4 所示.

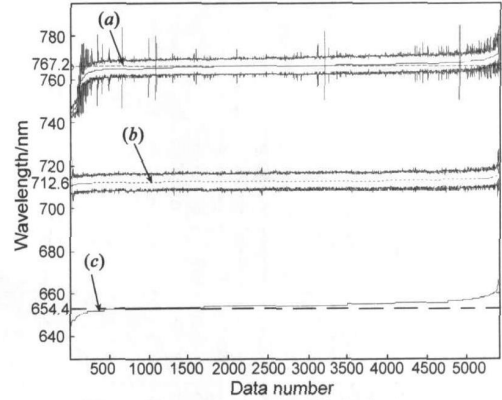


图 4 (a) $\lambda_{767.2}$ 位置处吸收检测结果
(b) $\lambda_{712.6}$ 位置处吸收检测结果
(c)零交叉点 x (图 3(c))检测结果

图 4 中, a, b 和 c 曲线都是升序排列,其中, a 和 b 曲线上分别画出了对应的 z_i 和 w_i 值.

如图 4 所示,本文方法可以直接将 $\lambda_{712.6}$ 和 $\lambda_{767.2}$ 处的两个吸收带较准确地检测出来.其中, $\lambda_{712.6}$ 处的吸收带检测结果较好,曲线 b 的均方差只有 0.55nm,这说明绝大多数的 M 型星都具有这一吸收带.另外,曲线 a 和 c 的均方差分别为 3.2714nm 和 2.0653nm.所以,本文检测吸收带位置的方法适用于 M 型星中的 $\lambda_{85.2}$ 、 $\lambda_{712.6}$ 和 $\lambda_{767.2}$ 处的吸收带.利用这三个吸收带的检测结果,制定约束关系如下: M 型星光谱中至少存在两个上述 TiO 吸收带.本文称该约束关系为“约束规则 2”.

利用“约束规则 1”和“约束规则 2”进行判别,就可以较好地识别 M 型星.

4 实验及结论

4.1 实验数据选择

实验中,选择了 SDSS DR2 中的 M 型星光谱 10816 条,其他类型天体有:类星体光谱 10000 条,恒星光谱 10000 条,星系光谱 10000 条.其中,恒星是指去除 M 型星的其他类型恒星光谱.将实验数据随机均分为两个数据集:数据集 1 和数据集 2.

4.2 阈值 σ 的确定

特征频率处幅值百分比的阈值 σ 由下面方法确定.选择数据集 1,利用“约束规则 1”和“约束规则 2”进行 M 型星判别,识别率和阈值 σ 有下面的关系:

如图 5 所示,在阈值 $\sigma < 0.32$ 时,整体识别率变化不大,为了更好地识别 M 型星,同时较好地地区分类星体,本文选择阈值 $\sigma = 0.3$.

4.3 实验结果及结论

数据集 1 的识别结果见表 1:

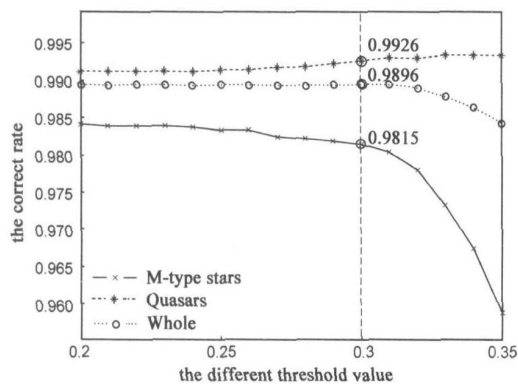
图5 不同阈值 σ 下的识别率

表1 数据集1的识别结果

天体名称	M 型星	类星体	恒星	星系	总体
正确率	98.15%	99.26%	98.74%	99.74%	98.96%

数据集2的测试结果见表2:

表2 数据集2的识别结果

天体名称	M 型星	类星体	恒星	星系	总体
正确率	97.56%	99.28%	98.64%	99.80%	98.79%

从表1和表2可以看出,各种天体的正确识别率都达到了97%以上,其中,类星体和星系的误识别率不到1%。

另外,测试本文识别方法对高红移类星体的正确鉴别情况。实验数据选择SDSS DR2中标注出的所有高红移类星体2215条,测试正确率为99.73%。可见本文方法可以较好地识别M型星,同时对类星体和星系的误识别率较低,实验结果令人满意。

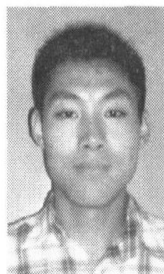
在巡天所得光谱中进行M型星识别是一项非常有意义但也是非常困难的工作。本文利用吸收带检测的方法对这个问题进行了研究,首先选取一定波长范围的光谱进行5层小波变换,对第5层小波系数进行频谱分析,得到M型星的特征频率,然后结合小波特征进行吸收带位置检测,依据检测结果设计决策规则,从而将M型星从众多类型天体中识别出来。大量真实光谱数据实验表明,本文方法十分有效,识别率高达97.56%。

参考文献:

- [1] Carlos Jaschek, Mercedes Jaschek. The Classification of Stars [M]. New York: Cambridge University Press, July 1987, 351-375.
- [2] M. J. Kurtz. Progress in Automation Techniques for MK Classification [D]. 1984.
- [3] 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.
Qin Dong mei, Hu Zharr yi, Zhao Yong heng. A PCA based efficient stellar spectra classification method [J]. Spectroscopy and Spectral Analysis, 2003, 23(1): 182-186. (in Chinese)

- [4] 黄凌云, 胡占义. 基于 Hough 变换的类星体红移测量方法 [J]. 光谱学与光谱分析, 2003, 23(1): 187-191.
HUANG Ling yun, HU Zharr yi. A new automatic quasars recognition technique based on PCA and hough transform [J]. Spectroscopy and Spectral Analysis, 2003, 23(1): 187-191. (in Chinese)
- [5] S Mallat. A theory for multiresolution signal decomposition: the wavelet representation [J]. IEEE Trans Pattern Anal Machine Intell, 1989, 11(7): 674-693.
- [6] 刘中田, 赵瑞珍, 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 Zhong tian, ZHAO Rui zhen 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. (in Chinese)
- [7] 刘中田, 吴福朝, 等. 基于卷积型小波包变换的谱线自动提取方法 [J]. 光谱学与光谱分析, 2006, 26(2): 372-376.
LIU Zhong tian, WU Fu chao et al. A method for auto extraction of spectral lines based on convolution type of wavelet packet transformation [J]. Spectroscopy and Spectral Analysis, 2006, 26(2): 372-376. (in Chinese)
- [8] 程佩清, 编. 数字信号处理教程 [M]. 北京: 清华大学出版社, 1998.
Cheng Pei qing. Digital Signal Processing Tutorial [M]. Beijing: Tsinghua University Press, 1998. (in Chinese)
- [9] 赵瑞珍, 胡占义. 谱线自动提取的小波变换零交叉点方法 [J]. 光谱学与光谱分析, 2005, 25(1): 153-156.
ZHAO Rui zhen, HU Zharr yi. Wavelet transform zero crossing method for auto extraction of spectral lines [J]. Spectroscopy and Spectral Analysis, 2005, 25(1): 153-156. (in Chinese)

作者简介:



刘中田 男, 1979年4月出生于安徽蚌埠, 2001年毕业于北京交通大学自动控制专业, 其后进入中科院自动化所模式识别国家重点实验室, 2006年7月获得工学博士学位, 现为北京交通大学电子信息工程学院教师, 从事模式识别、信号处理等方面的有关研究。

E-mail: liuzht@bjtu.edu.cn



李乡儒 男, 1972年5月出生于山东, 1993年进入山东师范大学数学系学习, 1997年获得理学学士学位, 2003年在吉林大学数学系获得理学硕士学位, 现在中科院自动化所模式识别国家重点实验室攻读博士学位, 研究方向为模式识别、机器学习等。

E-mail: xrl@nlpr.ia.ac.cn