

基于小波多尺度特征匹配的一类星体红移测量方法

刘中田¹, 李乡儒¹, 吴福朝¹, 赵永恒²

1. 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
2. 中国科学院国家天文台, 北京 100012

摘要 在中国正在实施的大型巡天项目(LAMOST项目)中,预计能获得 10^5 数量级的类星体光谱。文章旨在研究适用于LAMOST观测数据的类星体红移测量方法。为了克服信噪比较低的不利因素,文章采用小波变换的方法对类星体宽发射线进行特征提取,然后利用多尺度特征匹配的方法进行类星体红移测量。通过对sloan digital sky survey(SDSS) data release 2(DR2)中的15,715条类星体光谱的实验表明,在误差为0.02的范围内所用方法的正确率达到95.13%。该方法可对相对定标的类星体光谱数据进行红移测量,符合LAMOST数据的要求,可为天文学家进行类星体和宇宙大尺度等研究提供帮助。

主题词 小波变换; 特征提取; 特征匹配; 宽发射线

中图分类号: TN911.7 **文献标识码**: A **文章编号**: 1000-0593(2006)09-1738-04

引言

类星体是目前可观测到的光度最高、距离最远的天体,对宇宙大尺度结构研究具有重要价值。红移是类星体重要的物理参量^[1,2]。在我国正在实施的LAMOST项目中,预计能获得 10^5 数量级的类星体光谱数据。对于如此大量的巡天观测数据人工无法完成红移测量的任务,自动的光谱分析方法将成为必然的选择。

SDSS是一个比较成功的大型巡天计划,它对观测天体进行成像巡天和光谱巡天。在类星体红移测量方面,SDSS首先利用成像巡天数据测得光度红移值^[3],然后结合光谱巡天数据利用交叉相关和谱线证认等方法测得较准确的红移值。由于LAMOST只进行光谱巡天,只能利用类星体光谱数据进行红移测量,因而SDSS的红移测量方法不能用于LAMOST。类星体光谱数据的红移分布范围较大,信噪比较低,因而红移的自动测量较为困难。目前,研究类星体红移自动测量的文献很少,而且基本上都是基于谱线匹配的方法,如Hough变换法^[1]、伪三角法^[4]等。类星体光谱中一般具有比较宽的发射线,低信噪比使得宽发射线一般会叠加很多噪声,从而导致正确提取谱线的困难。因此,基于谱线匹配的方法有时很难得出较准确的红移值。

针对LAMOST数据信噪比较低这一不利因素,本文提出了一种特征提取的方法,用来提取类星体的宽发射线特

征,从而避开了低信噪比下的谱线提取这一难点。为了克服匹配范围较大对相关匹配的不利影响,本文采取了小波多尺度特征匹配的方法,以减少发生误匹配。本文方法的优点在于:(1)特征提取方法抗噪声能力强,简单易行;(2)多尺度特征匹配的方法很大程度地提高了红移测量的正确率。

1 类星体光谱的多尺度特征提取

一般来说,类星体光谱信噪比比较低,这给红移测量增加了很大困难。本文利用小波变换多分辨分析的特点,对类星体光谱进行多尺度变换,通过变换后的小波系数,提取出宽发射线的特征。实验研究发现,类星体的宽发射线特征主要表现在小波系数的第4,5,6层。本文选用SDSS的类星体模板,如图1所示。通过对类星体模板的特征提取的分析,在红移较低时,宽发射线特征在小波变换的第4,5层表现明显,在红移较高时,宽发射线特征主要表现在小波变换的第5,6层^[5]。也就是说,随着类星体红移的增大,发射线宽度在增加,体现宽线特征的小波变换尺度也在增加。本文采用第4,5,6层小波系数进行类星体特征提取,考虑的红移范围是0~5。

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 层半高通小波系数的算子用 W_i 表示,而算子 W_i^{-1} 表示对第 i 层半高通小波系数进行小波反变换。

收稿日期: 2005-06-06, 修订日期: 2005-10-20

基金项目: “863”计划(2003AA133060)和LAMOST项目资助

作者简介: 刘中田, 1979年生, 中国科学院自动化研究所博士研究生

定义 2: 算子表示对信号中小于零的部分置零, 然后按下式(归一化)运算

$$f^N(n) = f(n) / \sqrt{f(n) \cdot (f(n))^T} \quad (1)$$

其中, 上标 N 表示归一化后的值, 上标 T 表示行向量的转置, 标号在下文中含义相同。

图 2 为利用算子 \tilde{W}_i^{-1} 和 G 进行多尺度特征提取的示意图(光谱 $f(n)$ 先进行了 6 层小波变换^[6], 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])。

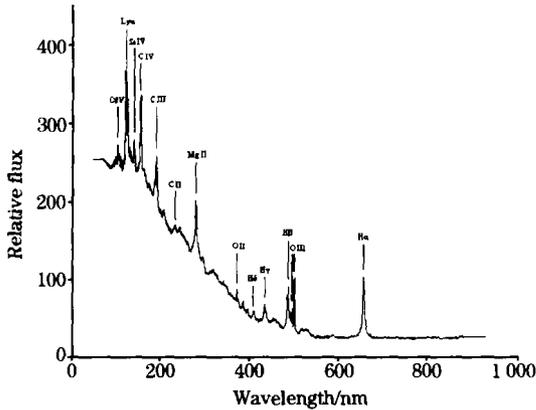


Fig 1 Quasar static template

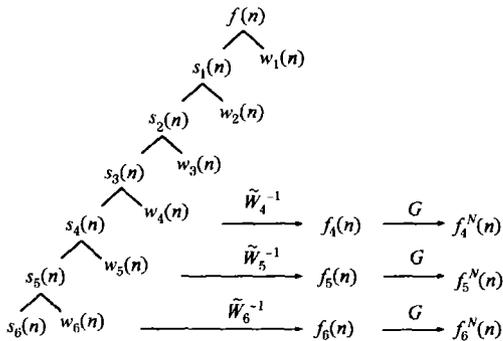


Fig 2 Wavelet transform and feature extraction

本文提取的多尺度特征可以表示为

$$f_i^N(n) = G\tilde{W}_i^{-1}W_i(f(n)), \quad i = 4, 5, 6 \quad (2)$$

其中 $i = 4, 5, 6$ 表示小波变换的第 4, 5, 6 层尺度。

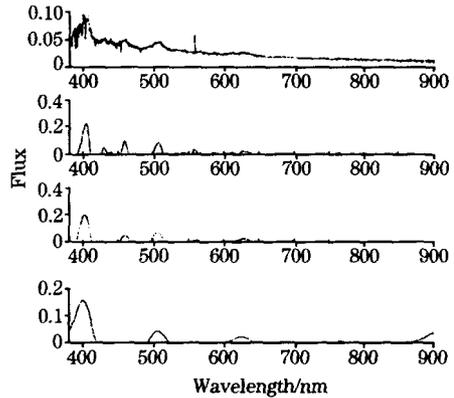
图 3 给出了 2 条实测类星体光谱及其多尺度特征。从图 3 可以看出: 本文方法较好地抑制了噪声的干扰, 得到了较为准确的宽谱线的特征。同时, 本文方法有效地去除了在 558.0 nm 附近的天光线。

2 类星体红移测量方法

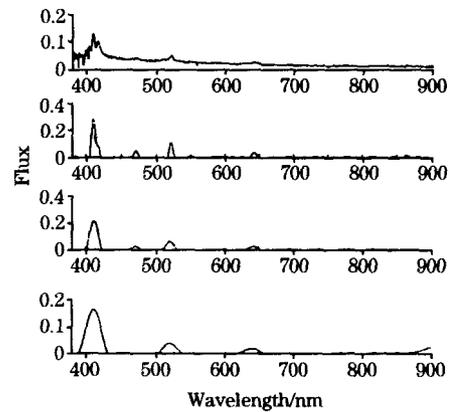
设类星体光谱信号为 $f(n)$, 类星体模板为 $q(m)$, m 是 n 离散的波长值。利用类星体模板, 结合类星体特征提取方法, 就可以进行类星体红移测量。下面详细介绍红移测量的几个主要步骤。

(1) 选取红移范围 0~5, 红移间隔 0.01, 得到红移值

$z_i, 1, 2, 3, \dots, 501$ 。求得模板光谱 $q(m)$ 的特征 $q_i^N(m) = G\tilde{W}_5^{-1}W_5(q(m))$, 对 $q_i^N(m)$ 按红移 z_i 用线性插值的方法求得 $q_i^*(n)$, 其中, $i = 1, 2, 3, \dots, 501$, n 表示波长范围和 $f(n)$ 相同。



(a) The multi-scaling feature extracted



(b) The multi-scaling feature extracted

Fig 3 The multi scaling feature extracted from quasar spectra

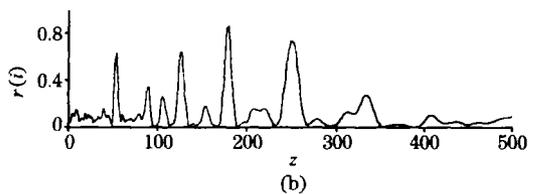
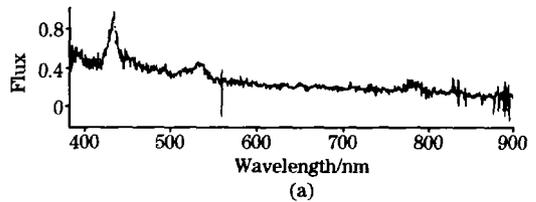


Fig 4 A quasar spectrum and the coefficient of correlation (a): A quasar spectrum; (b): The coefficient of correlation

(2) 求得类星体光谱 $f(n)$ 的特征 $f_i^N(n) = G\tilde{W}_5^{-1}W_5(f(n))$, 按下式求取相关系数 $r(i)$ ^[8]

$$r(i) = \sum_n f_i^N(n) \cdot q_i^*(n) \quad (3)$$

并选取 $r(i)$ 的前 8 个最大的极大值点对应的红移值作为红移候选值^[注1], 其中, i 对应离散化的红移值 $z_i, i = 1, 2, 3, \dots, 501$ 。图 4 表示一条类星体光谱及其对应的相关系数 $r(i)$ 。

(3) 利用类星体模板 $q(m)$ 按 8 个红移候选值用线性插值的方法得到 8 个近似光谱 $q_k(n), k = 1, 2, \dots, 8$, 其中 $q_k(n)$ 和 $f(n)$ 具有相同的波长范围和分辨率。根据公式(2) 得到多尺度特征

$$f_i^N(n) = G\tilde{W}_i^{-1}W_i(f(n)), i = 4, 5, 6 \quad (4)$$

$$q_{k,i}^N(n) = G\tilde{W}_i^{-1}W_i(q_k(n)), i = 4, 5, 6 \quad (5)$$

利用上面的特征, 根据下式得到多尺度特征匹配值 $Z(k)$

$$Z(k) = \sum_{i=4}^6 f_i^N(n) \cdot (q_{k,i}^N(n))^T, k = 1, 2, \dots, 8 \quad (6)$$

选择最大的 $Z(k)$ 对应的红移候选值, 作为所得的红移估计值 z_c ^[注2]。

(4) 在 z_c 的小区间上 $[z_c - 0.015, z_c + 0.015]$, 选取红移间隔 0.001, 利用多尺度特征匹配的方法进行红移求精运算, 求得较准确的类星体红移值 z_q 。

注 1: 通过对 $r(i)$ 研究发现, 若选取前 8 个最大的极大值点对应的红移值作为红移候选值, 则标准红移值在其中的可能性为 98.1%, 误差范围是 0.02。选取的红移候选值个数和相应的可能性如图 5 所示。

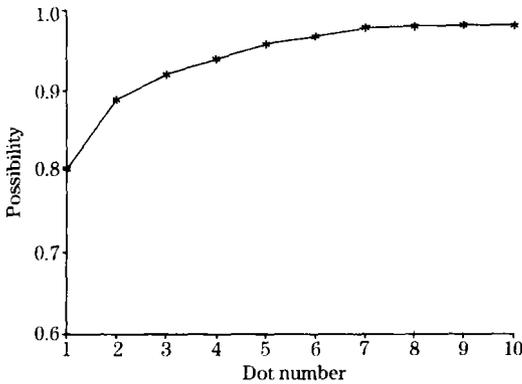


Fig 5 The number of the candidate of redshift and the possibility

从图中可以看出, 选择 8 个以上的红移候选值, 相应的可能性并没有明显地提高, 因此实验中选取了 8 个红移候选值。

注 2: 第 3 步的目的是从红移候选值中得到红移估计值 z_c 。设红移候选值为 $z_k^*, k = 1, 2, \dots, n$, 相应的多尺度特征匹配值为 $Z(k), k = 1, 2, \dots, n$, 判决规则如下

$$z_c = z_i^*, \text{ if } Z(i) = \max(Z(k), k = 1, 2, \dots, n) \quad (7)$$

其中, \max 表示向量中的最大值, $Z(i)$ 表示 $Z(k)$ 中的第 i 个值, z_i^* 表示 z_k^* 中和 $Z(i)$ 对应的第 i 个值。本文把这种判定红移值的方法称为多尺度特征匹配的方法。

3 实验与结论

3.1 数据选择

本文实验数据是从 SDSS DR2 中随机选取的 15, 715 条类星体光谱数据, 红移值以 SDSS 的给出值作为参考标准。

3.2 实验结果及分析

利用本文方法对实验数据进行红移值测量, 选取适当的红移误差范围对结果进行统计, 所得的正确率如表 1 所示。

Table 1 The correct rate of redshifts obtained by our method

红移误差范围	0.01	0.02	0.03	0.05
正确率/%	90.93	95.13	95.90	96.32

把本文测得的红移值和 SDSS 给出的红移值作比较, 结果如图 6 所示。

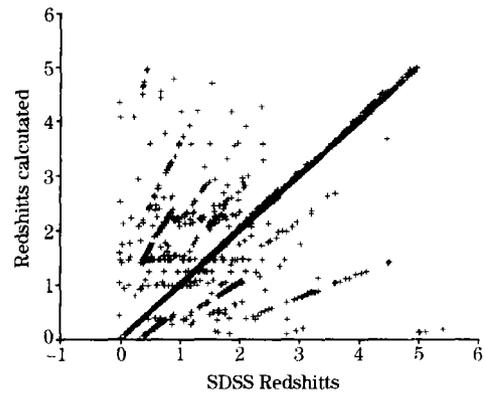


Fig 6 The redshift result

在上图中, 横坐标是 SDSS 给出的红移参考值, 纵坐标是本文测得的红移值, 每一个“+”代表一个对应点。

从图 6 可以看出: 类星体红移值较大时, 测量结果的误差一般也较大。利用相对误差范围进行统计, 结果如表 2 所示。

Table 2 The correct rate with relative error

相对误差范围	0.01	0.02	0.03
正确率/%	95.16	96.32	96.63

相对误差计算公式

$$\sigma = \frac{|z_q - z_{sdss}|}{z_{sdss}} \quad (8)$$

其中, z_q 是本文方法求得的红移值, z_{sdss} 是 SDSS 给出的红移值。

3.3 结论

实验结果表明, 本文方法仅利用类星体光谱信息, 就可以较准确地测得红移值, 而且实验正确率较高, 基本上满足 LAMOST 对类星体红移测量的要求。由于采用了相关运算, 该方法的计算量较大。下一步, 我们将尝试使用基于机器学习的方法进行类星体红移测量, 以提高运算速度和测量正确率。

参 考 文 献

- [1] HUANG Ling yun, HU Zhanyi(黄凌云, 胡占义). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(1): 187.
- [2] LIU Rong, DUAN Furqing, LUO Ali(刘蓉, 段福庆, 罗阿里). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(7): 1155.
- [3] Michael A Weinstein, Gordon T Richards, et al. Astrophys. J. Suppl., 2004, 155: 243.
- [4] QIU Bo, HU Zhanyi, ZHAO Yongheng(邱波, 胡占义, 赵永恒). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2002, 22(4): 695.
- [5] Darren S Madgwick, Paul C Hewett, et al. Monthly Notices of the Royal Astronomical Society, Volume 334, Issue 1, 2002. 209.
- [6] LIU Zhongtian, ZHAO Ruizhen, ZHAO Yongheng, et al(刘中田, 赵瑞珍, 赵永恒, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(7): 1158.
- [7] QIN Taigui, ZHANG Xueying, et al(覃太贵, 张学英, 等). Journal of Mathematics(数学杂志), 2004, 24(3): 323.
- [8] XU Xin, LUO Ali, WU Furchao, et al(许馨, 罗阿里, 吴福朝, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(6): 996.

A Method for Obtaining Redshifts of Quasars Based on Wavelet Multi-Scaling Feature Matching

LIU Zhongtian¹, LI Xiangru¹, WU Furchao¹, ZHAO Yongheng²

1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
2. National Astronomical Observatory, Chinese Academy of Sciences, Beijing 100012, China

Abstract The LAMOST project, the world's largest sky survey project being implemented in China, is expected to obtain 10^5 quasar spectra. The main objective of the present article is to explore methods that can be used to estimate the redshifts of quasar spectra from LAMOST. Firstly, the features of the broad emission lines are extracted from the quasar spectra to overcome the disadvantage of low signal to noise ratio. Then the redshifts of quasar spectra can be estimated by using the multi-scaling feature matching. The experiment with the 15, 715 quasars from the SDSS DR2 shows that the correct rate of redshift estimated by the method is 95.13% within an error range of 0.02. This method was designed to obtain the redshifts of quasar spectra with relative flux and a low signal to noise ratio, which is applicable to the LAMOST data and helps to study quasars and the large scale structure of the universe etc.

Keywords Wavelet transform; Feature extraction; Feature matching; Broad emission line

(Received Jun. 6, 2005; accepted Oct. 20, 2005)