



论文

从海量星系光谱中搜索超新星候选光谱

屠良平^{①②③*}, 罗阿理^③, 吴福朝^①, 赵永恒^③

① 中国科学院自动化研究所, 北京 100190;

② 辽宁科技大学理学院, 鞍山 114051;

③ 中国科学院光学天文重点实验室(国家天文台), 北京 100012

* E-mail: lptu@nlpr.ia.ac.cn; tuliangping@tom.com

收稿日期: 2009-09-16; 接受日期: 2010-03-17

国家自然科学基金资助项目(批准号: 60773040, 10973021)

摘要 完善了一种在海量星系光谱中快速搜索超新星光谱的方法. 该方法主要包括五个步骤, 第一, 光谱的预处理, 包括去噪、退红移等处理过程; 第二, 对每一条光谱进行星系成分和超新星成分的快速分解, 并计算本文定义的超新星统计特征描述向量; 第三, 样本约减, 即利用基于局部孤立性因子的离群搜索算法对海量光谱数据样本进行约减; 第四, 对约减后的结果与所有模板进行交叉相关匹配, 根据匹配结果及相应参数获得初步的超新星候选光谱; 第五, 人工排除, 即根据超新星光谱明显的谱线特征排除一些误匹配结果, 最终得到特征相对明显的超新星候选光谱. 根据上述方法, 对选自美国斯隆数字巡天第七期释放数据(SDSS-DR7)的具有较高信噪比的 294843 条星系光谱进行搜索, 获得了 36 条含爆发超新星的星系光谱候选, 其中 9 个为本文首次发现并报告, 15 个已以快报形式被本文作者公开报告, 另外 12 个为已发现并公开命名的超新星光谱. 根据最匹配模板的类型, 24 个未命名的超新星候选光谱包括 20 个 Ia 型, 3 个 Ic 型和 1 个 II 型样本.

关键词 超新星, 光谱分解, 样本约简, 局部孤立性因子, 交叉相关匹配

PACS: 42.30.Sy, 95.55.Qf, 95.85.Kr, 42.30.Tz, 98.62.Py

超新星(Supernovae, SNe)是某些大质量恒星在演化末期经历的一种剧烈爆炸, 这种爆炸极其明亮, 可以照亮整个宿主星系, 过程中所发出的能量几乎相当于太阳穷其一生所发射能量的总和. 超新星是目前已知的最剧烈的天文现象之一, 其相关研究也是当今天体物理学研究最活跃的核心领域之一, 它不仅是恒星演化理论发展的一个极重要的环节, 而且是天体演化学和极端条件下物理学的交叉点. 在宇宙学参数测量、暗能量探索研究方面具有十分重要的作用, 对超新星的相关研究几乎与所有天体物理

前沿领域密切相关^[1].

超新星的研究必须要有超新星的实时观测样本, 为此需要开展超新星的巡天工作, 相应工作目前基本可以分为两类: (i) 天文爱好者用小型望远镜通过监视较亮的近邻星系搜寻超新星, 他们的发现多为较亮的超新星; (ii) 专业天文学家所作的超新星巡天工作, 由于超新星研究在宇宙学参数测量和暗能量研究的巨大作用以及光学和计算机技术的极大发展, 自 20 世纪 90 年代以来国际上开始实施许多专门的大型的超新星巡天计划, 而且许多都是多国家

多机构一起合作, 如: Sloan Digital Sky Survey(SDSS): Supernova Survey^[2,3], High-Redshift Type Ia Supernova Survey with HST^[4], Lick Observatory Supernova Search(LOSS)^[5], Carnegie Supernova Project (CSP)^[6]和Nearby Supernova Factory (SNfactory)等^[7], 另外一些以寻找和跟踪变源为目的的大型巡天计划对超新星搜寻的贡献也很大, 如 2009 年底已有超新星发现的大型巡天计划Pan-STARRS^[8] (<http://pan-starrs.ifa.hawaii.edu/>), 据估计它将成为近期搜寻超新星的最大贡献者.

这些大型巡天计划基本上采取搜索手段是一样的, 就是通过对同一天区进行不断的光学观测和图像对比, 搜寻亮度变化的目标进行跟踪观测, 然后判断其是否是超新星以及相应的类型. 他们有一个共同的特点是要有一个专门的观测能力强大的望远镜作为基础, 然后多个其他望远镜作辅助定点观测, 这一特点限制了类似巡天计划在国内的开展.

我们要想在这方面有所突破, 只能依靠现有的设备进行不同于他们的计划. 本文即是基于这一出发点, 为将来利用国家大科学工程Large Sky Area Multi-Object Fiber Spectroscopic Telescope(LAMOST, <http://www.lamost.org/>)有效搜寻超新星做准备, 研究完善了一种可以实时地有效地从大量星系光谱中搜寻超新星候选的方法. 结合模式识别方法对大量数据进行统计意义的特征描述, 本文获得了这一不用测光和图像对比的有效搜寻超新星候选(因为没有更多的光谱证认, 这里所指的超新星本文都暂称超新星候选)方法, 该方法通过预处理、星系成分扣除后计算星系光谱的超新星统计特征描述向量在数据中的孤立性进行离群点搜索, 然后通过Blondin和Tonry的超新星光谱证认方法SNID(Supernovae Identification)^[9,10]进行模板初步证认, 最后通过引入一些评估参数获得超新星候选. 因为SNID中使用的模板是有限的实测模板, 而且待测光谱会受到天光污染和其他噪声的影响, 所以最后获得的数量已经很少的超新星候选可通过人工进一步筛查.

本文利用该方法对SDSS的具有较高信噪比的294843条光谱数据进行分批搜索, 获得了36个超新星候选光谱, 其中包括12个其他文献已有记录并证认的, 24个本文作者发现的(其中的15个已以快报形式公布^[11], 9个本文首次公布). 结果证明了本文的方法是可行和有效的. 本文方法在将来应用时所能体

现的优点是不用专门使用LAMOST进行预定观测, 只要对每天根据其他团组研究计划所获得的大量星系光谱进行实时的自动分析处理, 搜寻获得超新星候选光谱. 相比于其他大型巡天计划巨大的资源消耗, 本文方法所需要利用的资源几乎可以忽略. 而且通过对LAMOST观测能力的简单估计以及本文在SDSS光谱数据中发现的超新星候选数量, 预期利用LAMOST的搜索数量也会较以前国内原北京天文台^[12,13]和鹿林天文台的超新星巡天计划有较大提高.

1 方法

1.1 数据约减

由于获得的光谱数量非常大, 超新星爆发高亮度持续时间相对短暂, 如果要应用于实时的搜索超新星计划, 则首先要能够快速地缩小候选范围. 屠良平等^[14]曾提出一种超新星候选范围约减的算法, 在该算法思想的基础上进行了改进并用于本文的数据约减步骤, 这一约简算法可以将搜索范围缩小到原来的1%~2%. 该方法分为: (i) 星系超新星混合光谱分解; (ii) 孤立性度量计算, 孤立性较大的光谱即为待选光谱. 该方法部分思想在文献[11]曾有提及, 本文增加了光谱分解内容并详细介绍如下.

(i) 星系超新星混合光谱快速分解和超新星统计特征描述. 一般地, 有超新星爆发的星系的光谱 D 基本由寄主星系光谱成分 G , 超新星光谱成分 SN 和噪声 σ 构成, 即

$$D = G + SN + \sigma.$$

这里, 噪声可能包括测量误差, 天光污染等. 对混合光谱进行分解的目的是提取 D 中超新星光谱成分 SN 并去除星系光谱成分和噪声. 本文中要对所有的光谱进行分解, 并且利用的是分解后 SN 结果(对没有超新星爆发的星系光谱分解后结果是无实际意义的向量). 为了消除噪声的影响, 这里首先要进行去噪声处理, 本文采取了小波变换的方法进行去噪处理^[15,16], 去噪效果见图1.

在已有的超新星的寄主星系成分扣除或称为星系超新星成分分解方法中, 主要有两种算法, 一种是Madgwick等人^[17]利用基于星系PCA(Principal Component Analysis, 主成分分析)^[18]特征谱的星系成分重

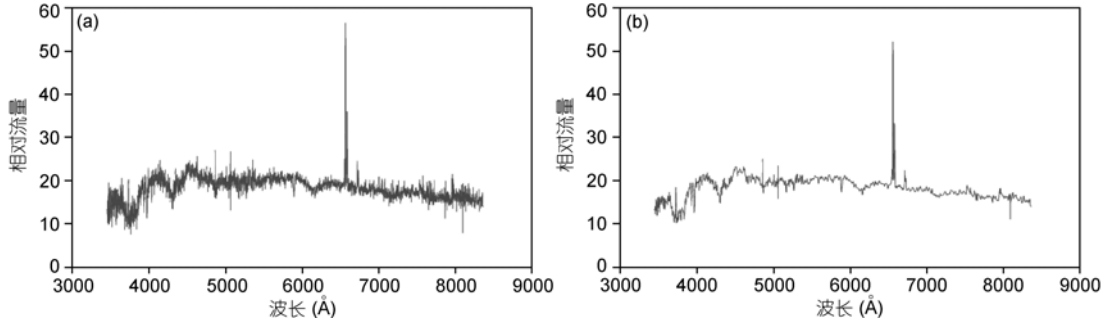


图 1 光谱去噪示例

(a) 原始光谱; (b) 去噪后光谱

构分解方法, 该文对SDSS-DR1(SDSS: Data Release 1)^[19]中的约 11 万条星系光谱做了星系成分扣除后进行超新星搜索, 获得了 19 条含超新星爆发的星系光谱. 另外一种是基于 χ^2 模板拟合技术的星系成分重构分解方法^[20-23], Zheng等人^[24]将其改进后用于对已发现的超新星观测光谱进行星系成分扣除, 进一步较准确地确定了超新星的年龄, 类型等信息. 直接基于星系PCA特征谱的星系成分扣除最简单易行, 该方法选择一批纯星系光谱样本进行PCA分析, 获取其方差贡献率较大的前 m 条特征光谱 $g_1, g_2, \dots, g_m$ 作为一组基, 每条混合光谱 D 的星系成分光谱 $G_{\text{reconstruction}}$ 为光谱 D 在该组基构成的子空间上的投影向量, 即

$$G_{\text{reconstruction}} = \sum_{i=1}^m a_i g_i,$$

其中

$$a_i = \langle D, g_i \rangle, \quad i = 1, 2, \dots, m.$$

利用上式可得残差谱 $S_{\text{residual}} = D - \sum_{i=1}^m a_i g_i$, 即完成星系成分扣除. 此方法优点是简单快速易行, 缺点是其分解效果依赖于一个隐含条件, 即超新星模板子空间与星系模板子空间应正交, 显然这一条件通常不会满足, 所以这种方法只是在超新星成分占主导时残差谱能基本保持超新星光谱的谱线形状特征, 从而对后继基于超新星光谱谱线形状特征的分析不产生重大影响.

Howell等人^[20,21]用 χ^2 模板拟合技术进行星系成分扣除. 该方法利用一个引入的 χ^2 匹配算法计算星系模板库和超新星模板库中每一对模板线性组合逼近混合光谱的程度, 进而选择最佳的模板对的线性组合完成混合光谱的分解和星系成分扣除, 即

$$\chi^2 = \sum_{\lambda} [D(\lambda) - aG_k(\lambda) - bSN_l(\lambda)]^2, \quad (1)$$

这里 λ 为波长, a 和 b 为待定系数, 对每一星系超新星模板对 G_k 和 SN_l , 寻找最优的 a', b' 使得 χ^2 最小, 然后选择所有的星系超新星模板对对应最优 χ^2 值中最小值对应的分解结果为最终结果:

$$(a^*, b^*, G_{k^*}, SN_{l^*}) = \arg \min_{\text{all } G_k, SN_l, a', b'} \{\chi^2\},$$

$G_{\text{reconstruction}} = a^* G_{k^*}$ 和 $S_{\text{residual}} = D - a^* G_{k^*}$ 即为重构的星系成分光谱和超新星成分光谱. 相应地, 如果将星系模板用PCA特征谱 $g_i (i = 1, 2, \dots, m)$ 代替, 则(1)式可写为

$$\chi^2 = \sum_{\lambda} [D(\lambda) - \sum_{i=1}^m a_i g_i(\lambda) - bSN_l(\lambda)]^2. \quad (2)$$

这一方法在理论和实际上有较好的解释和效果, 所以多数的超新星后继光谱观测中的星系成分扣除采用此方法, 如Lidman等人^[22], Hook等人^[23]和Zheng等人^[24].

上述两种方法各有优缺点, 第一种虽然简单但分解结果较差, 第二种可充分考虑各种因素, 采取的近似逼近较好的解释了混合光谱的合成, 但由于对每一个超新星模板都要进行匹配分解计算, 所以时间消耗较大, 不适合用于大规模光谱数据处理. 为了克服上述缺点, 本文将(2)式中的超新星模板 SN 用超新星模板空间的PCA特征谱的线性组合来替换, 本文称之为基于二元混合PCA特征光谱的星系超新星成分分解方法. 本文中使用选定星系模板PCA特征谱和超新星模板PCA特征谱作为基向量, 并改进了相应的计算过程, 使得该方法在保持 χ^2 模板拟合方法的效果下大大缩减了解析时间消耗, 进而可以用于实时搜寻超新星光谱计划. 下面详细介绍本文的

方法:

设已去除噪声的混合光谱 $D = G + S$, G 表示星系光谱, S 表示超新星光谱. 记 g_i, s_j ($j=1, 2, \dots, n$, $i=1, 2, \dots, m$) 分别为星系模板库的前 m 个 PCA 特征谱和超新星模板库的前 n 个 PCA 特征谱. 则有

$$\begin{aligned} G &\approx a_1 g_1 + a_2 g_2 + \dots + a_m g_m, \\ S &\approx a_{m+1} s_1 + a_{m+2} s_2 + \dots + a_{m+n} s_n. \end{aligned} \quad (3)$$

因此, $D \approx \sum_{i=1}^m a_i g_i + \sum_{j=1}^n a_{m+j} s_j$.

因为 g_i 和 s_j 通常并不是正交的, 所以不能直接利用 D 与每个 g_i, s_j 的内积来计算系数 a_i , 事实上前文第一种分解方法就是默认每一 g_i 和 s_j 正交来得到分解结果的, 而导致实验结果较差. 本文使用线性空间基之间的线性变换来计算上式中的所有组合系数. 不妨记 $\bar{D} = \sum_{i=1}^m a_i g_i + \sum_{j=1}^n a_{m+j} s_j$, 向量组 $g_1, g_2, \dots, g_m, s_1, s_2, \dots, s_n$ 通常是线性无关的, 所以对任意固定的光谱 $\bar{D}$, 其对应系数 a_i ($i=1, 2, \dots, m, m+1, \dots, m+n$) 是唯一的. 令 $A = (g_1, g_2, \dots, g_m, s_1, s_2, \dots, s_n)$, 它为 $d \times (m+n)$ 矩阵, d 为各向量维数, 则有 $\bar{D} = A(a_1, a_2, \dots, a_{m+n})^T$. 对 A 进行奇异值(SVD)分解, 存在 d 阶正交矩阵 U 和 $m+n$ 阶正交矩阵 V 使得 $A = U \begin{pmatrix} T_{m+n} \\ 0 \end{pmatrix} V^T$. 于是, 令 $U_1 = U(:, 1:m+n)$, 则 $A = U_1 T_{m+n} V^T$, 且 $U_1^T U_1 = I_{m+n}$ (I_{m+n} 为 $m+n$ 阶单位矩阵), 因此

$$\bar{D} = A(a_1, a_2, \dots, a_{m+n})^T = U_1 T_{m+n} V^T (a_1, a_2, \dots, a_{m+n})^T$$

$$\begin{aligned} \Rightarrow U_1^T \bar{D} &= T_{m+n} V^T (a_1, a_2, \dots, a_{m+n})^T \\ \Rightarrow (a_1, a_2, \dots, a_{m+n})^T &= (T_{m+n} V^T)^{-1} U_1^T \bar{D}, \end{aligned}$$

近似地, 有

$$(a_1, a_2, \dots, a_{m+n})^T = (T_{m+n} V^T)^{-1} U_1^T \bar{D},$$

再根据(1)式即可得星系成分及其含超新星成分的残差谱.

在进行 SVD 分解计算时, 因为光谱维数较高而导致有相当大冗余计算, 对海量的光谱数据处理而言, 时间消耗不能满足实时性的要求, 所以本文使用另一种更直接的计算效率更高的方法来计算所有系数 $a_1, a_2, \dots, a_m, a_{m+1}, \dots, a_{m+n}$.

通过对向量组 $g_1, g_2, \dots, g_m, s_1, s_2, \dots, s_n$ 标准正交化得到标准正交向量组 $\beta_1, \beta_2, \dots, \beta_m, \beta_{m+1}, \dots, \beta_{m+n}$, 且存在 $m+n$ 阶方阵 B 满足

$$(\beta_1, \beta_2, \dots, \beta_m, \beta_{m+1}, \dots, \beta_{m+n}) = AB.$$

故原系数可通过下式计算

$$(a_1, a_2, \dots, a_{m+n})^T = B(b_1, b_2, \dots, b_m, b_{m+1}, \dots, b_{m+n})^T,$$

这里, $b_i = \langle \bar{D}, \beta_i \rangle$ (实际计算中用 $b_i = \langle D, \beta_i \rangle$ 代替), ($i=1, 2, \dots, m+n$). 根据所得系数及(3)式即可得到近似分解结果. 图 2 是含超新星成分的星系光谱分解示例, 原始光谱都进行了归一化, 并且实测光谱进行了去噪.

(ii) 孤立性计算及数据约减. 由于含超新星爆发的星系光谱相对于一般的星系光谱而言非常稀少, 所以在大量样本集中超新星候选光谱应具有强烈的离群性质, 相应的每一条星系光谱经过上述的分解过

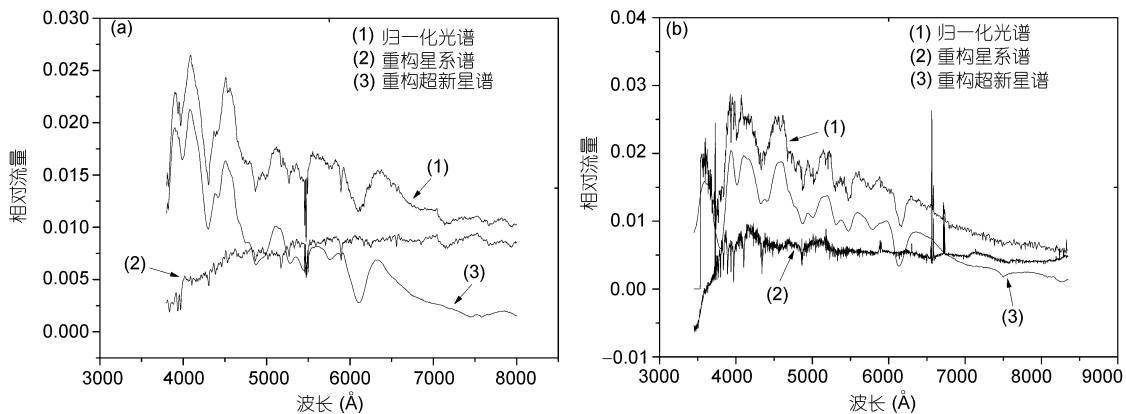


图 2 星系超新星光谱分解示例

(a) 模拟光谱分解; (b) 实测光谱分解

程得到的超新星成分光谱 $S' = a_{m+1}s_1 + a_{m+2}s_2 + \dots + a_{m+n}s_n$ 也保持了该性质, 即纯星系光谱的分解超新星成分(一个无意义的向量)应该也保留星系的聚类性质, 而含超新星爆发的星系光谱分解得到的超新星成分光谱也当然具有强烈的离群性. 不同于文献[14], 本文提出了基于分解后的超新星成分光谱的孤立性度量计算来进行数据约减的方法, 并且仍然采用Breunig等人^[25]提出的 k 局部孤立性因子(Local Outlier Factor, LOF)来度量样本光谱在样本集中的孤立性大小, 相关概念可参考原文献. 本文中距离采用欧氏距离, 在欧氏距离下, 可证明下面定理成立.

定理 1 每一分解超新星成分光谱 $S' = a_{m+1}s_1 + a_{m+2}s_2 + \dots + a_{m+n}s_n$ 在所有分解结果集合中的局部孤立性因子等于系数向量 $(a_{m+1}, \dots, a_{m+n})$ 在所有分解结果系数向量集合中的局部孤立性因子, 即

$$LOF_k(S') = LOF_k((a_{m+1}, \dots, a_{m+n})).$$

证明 记 $S'_1 = a_{1,m+1}s_1 + a_{1,m+2}s_2 + \dots + a_{1,m+n}s_n$, $S'_2 = a_{2,m+1}s_1 + a_{2,m+2}s_2 + \dots + a_{2,m+n}s_n$ 分别表示两个分解的超新星成分光谱, 因为超新星特征光谱向量组 $\{s_1, s_2, \dots, s_n\}$ 是标准正交向量组, 则二者距离

$$d(S'_1, S'_2) = d((a_{1,m+1}, a_{1,m+2}, \dots, a_{1,m+n}), (a_{2,m+1}, a_{2,m+2}, \dots, a_{2,m+n})),$$

而局部孤立性因子的计算全部是基于此距离的, 所以结论成立.

因为本文中分解是对所有星系光谱进行的, 而含超新星成分的只是少数, 所以本文称某个光谱的超新星成分分解结果 $S' = a_{m+1}s_1 + a_{m+2}s_2 + \dots + a_{m+n}s_n$ 对应的系数向量 $a' = (a_{m+1}, a_{m+2}, \dots, a_{m+n})$ 为原星系光谱的超新星统计特征描述向量(Supernovae Statistical Characterization Vector, SNSCV). 相应星系光谱在样本集中的孤立性度量就用其超新星统计特征描述向量的孤立性度量来代替, 这样可以大大缩减计算时间消耗. 本文数据约减算法步骤如下:

- (i) 计算 Ia 型超新星特征子空间的一组基向量;
- (ii) 预处理所有待选光谱, 包括去噪, 退红移, 及流量归一化;
- (iii) 对每一条星系光谱进行星系超新星光谱分解, 并获得该光谱的超新星特征描述向量 a' ;
- (iv) 计算所有样本对象 p 在样本集 D 中的 k (本文 $k=5$) 距离 k -distance 和 k 距离邻域, 进而计算该

对象 p 的局部可达密度;

(v) 根据 p 的局部可达密度计算其 k 局部孤立性因子, 然后对所有 D 中对象的 k 局部孤立性因子由大到小排序, 最终选取一定比例(比例可根据实际需求选择, 本文实验中取 1%比例)的局部孤立性较大的样本为超新星候选体后继搜寻前的选择范围.

1.2 基于 SNID 技术的超新星候选筛选

本文采用基于交叉相关匹配的方法对缩减后的样本进行继续筛选, 该方法由Blondin 和 Tonry^[9,10]基于Tonry等人^[26]的交叉相关方法测量星系红移演化而来, 其目的是利用模板库中的所有模板与目标光谱在不同红移处的相关性大小来确定超新星光谱的类型(一般包括有Ia, Ibc, II型及相应的子型)、年龄(本文超新星年龄指的是距离光极大时的天数)和红移, 称作Supernova Identification (SNID). 在本文中用该方法得出的结果可用来作为进一步筛选超新星光谱时的参考, 方法介绍如下.

交叉相关技术 所谓交叉相关就是将一条红移为 z_s 的光谱 $s(n)$ 与一条红移为零已知类型年龄的模板光谱 $t(n)$ 在不同的红移步长下做相关, 即确定在 $(1+z_s')$ 的波长刻度下相关函数 $c(n)=s(n)*t(n)$ 最大, (“*”表示交叉相关乘积), 则相应最佳匹配就是具有红移 z_s' 的模板 $t(n, z_s')$. 计算中一般使用对数波长, 此时, 红移相当于对数波长的平移. 本方法只采用了 SNID 中的一部分, 并根据数据加入了其他的一些处理, 主要步骤为退红移, 强窄线去除(超新星不具有的特征), 对数波长变换, 去除伪连续谱, 滤波平滑和最后交叉相关匹配. 本文采用 SNID 的光谱模板库, 该模板库包括 94 个超新星的不同年龄光谱和其他 17 条非超新星光谱模板, 总数为 1515 条. 超新星光谱来源有 SUSPECT 公开数据 (<http://bruford.nhn.ou.edu/~suspect/index1.html>) 和 CFA 超新星计划的一部分 (<http://www.cfa.harvard.edu/supernova/SNarchive.html>). 具体流程如下:

- (i) 退红移及利用小波变换去除强窄发射线;
- (ii) 波长变换: 将波长范围 $[l_0, l_1]$ 通过对数项的线性变换变换为 N 个坐标点, $n = A \ln l_{n,n} + B$, 其中 $A = N / \ln(l_1 / l_0)$, $B = -N \ln l_0 / \ln(l_1 / l_0)$, l_0, l_1, N 分别固定取 2500, 10000, 1024. 对应可计算每一个点 n 的原始波长坐标 $l_{n,n} = l_0 e^{n \times dl_n}$, 其中 $dl_n = \ln(l_0 / l_1) / N$;

(iii) 伪连续谱去除: 本文采用四次多项式拟合伪连续谱, 光谱与伪连续谱相除再减 1 即去除连续谱, 然后再将光谱零均值化, 即加减一个常数使得流量的平均值为零;

(iv) 光谱平滑: 本文采用均值滤波对光谱进行平滑处理;

(v) 交叉相关匹配: 通过与每一个模板进行交叉相关计算获得目标光谱与每一个模板的最大相关值、最大相关红移等参数;

(vi) 对所有模板按相关值(该相关值包含有光谱重叠因子, 即待测光谱波长范围和模板光谱波长范围比例)大小降序排列, 按阈值 T 截取, 分别计算最大相关值大于 T 的模板数 M , 其中的超新星模板数 SN_m 以及其中的 Ia 型超新星模板数 $IaSN_m$. 然后根据最相关模板是否为 SN 以及 $SN_fraction = SN_m/M$ 和 $Ia_fraction = IaSN_m/M$ 的值筛选可能的超新星候选;

(vii) 对(vi)中所得结果再通过人眼观察谱线特征确定超新星候选.

图3给出了一条光谱在交叉相关匹配前的预处理

理过程图示. 图中右边子图中横轴意义是对原始波长 W 的一个对数项线性变换 $A \ln W + B$, A 和 B 是两个常数, 该变换可有利于计算红移影响, A 和 B 的计算具体见上述步骤 2.

2 实验

2.1 数据选择

因为LAMOST仍在试运行阶段, 暂时不能获得有效数据进行检验, 所以本文使用美国斯隆数字巡天第七期光谱释放数据SDSS-DR7(Sloan Digital Sky Survey, Data Release 7)^[27]来验证本文方法的有效性, 该数据共有 2000 多个天区(Sloan 称plate), 每个天区观测目标 640 个, 总共有约 120 万条光谱, 包括有恒星, 星系, 类星体等. 本文选择了其中红移小于 0.25 (事实上Sloan的星系红移绝大多数小于此数值), g,r,i 波段信噪比都大于 10 的星系光谱 294843 条, 从中搜索超新星光谱. 由于计算机内存原因, 在数据约减阶段本文分批次进行实验, 每次选择约 5000 条光谱按

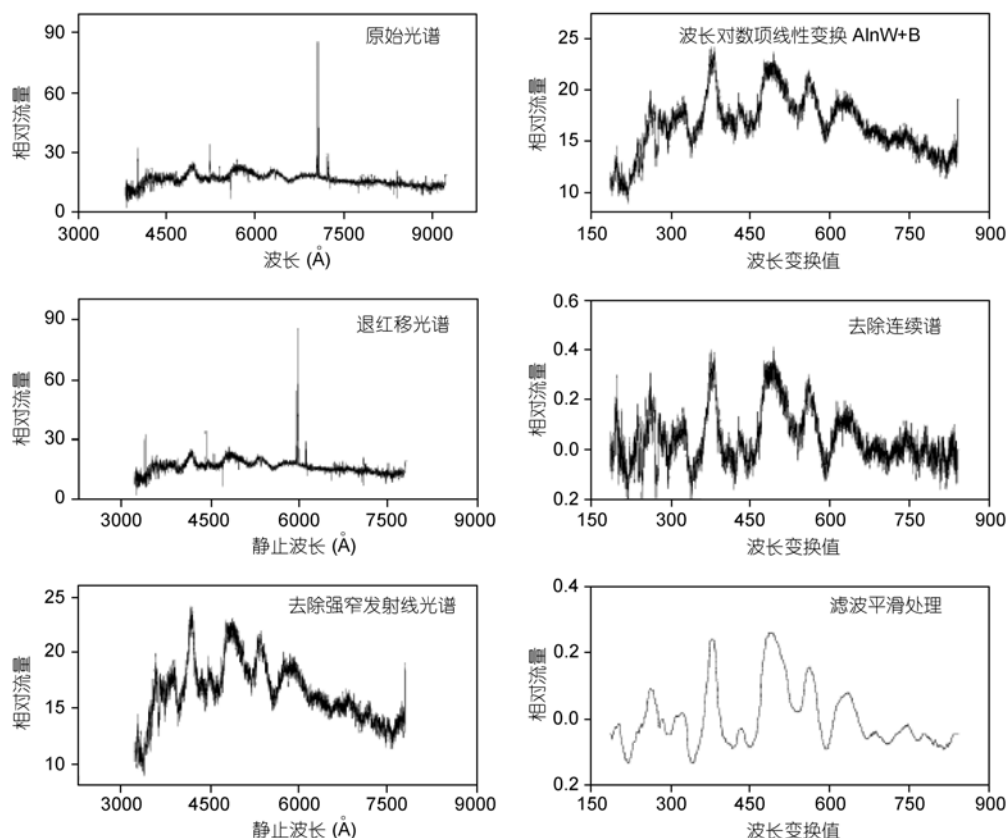


图3 交叉相关匹配前的光谱预处理示例

1%比例进行约减.

因为实验数据是SDSS的实测光谱, 所以本文中星系PCA特征谱选择Yip等人^[28]根据SDSS中约 17 万条星系光谱计算得到的 10 条PCA特征谱, 波长覆盖范围 345~834.87 nm, 点数为 3839. 超新星模板库采用Nugent^[29]的超新星理论模板库 (http://supernova.lbl.gov/~nugent/nugent_templates.html), 包含有多种类型对应不同年龄的超新星模板, 因为除Ia型以外的光谱模板数量还有待增加完善, 所以本文仅针对Ia型进行实验, 本文使用Ia-normal型超新星光谱 40 条, 距离V波段光极大时间分别为-9~30 days, 与星系模板波长保持一致, 所有超新星模板都插值到同一波长范围, PCA分析后取前12条PCA特征谱作为分解时的超新星特征谱, 方差贡献率大于 0.9999. 图 4 分别是星系和超新星的前两个PCA特征谱.

2.2 实验结果

通过数据约减共获得 2948 条光谱, 然后利用基于SNID技术的超新星光谱筛选方法获得约 150 条光谱. 通过人工筛选后得到了 36 个超新星候选光谱, 其中 12 个是已知的, 24 个为新发现的, 这 36 中的 25 个(10 个已知的和 15 个新发现的)本文作者已经在文献[11]中以快报形式报告, 另外 11 个中的 9 个为本文首次发现, 2 个为已知的. 图 5 为新发现的 24 个超新星候选光谱, 表 1 列出了所有发现的超新星候选光谱信息.

2.3 实验结果分析

实验的结果中既有之前已经证认的超新星, 也有新发现的超新星候选光谱, 它们中大多数有较明显的超新星光谱特征, 这表明本文利用光谱来搜索超新星候选的方法是行之有效的. 另外由于实验光谱数据对应目标已经不能进行后继跟踪观测证认, 还有一些特征不是十分明显的可疑光谱不能确认而没有列出, 可以预见, 在将来的 LAMOST 实时观测中, 超新星的发现效率会有所增加. 另外在数据约减后的光谱数据中约有一半左右是具有残谱的光谱, 即数据不完整, 这也说明了本文采用的离群搜索方法确实能提取出特殊的光谱数据. 需要说明的是最终的光谱数据中有一个是明显的 II 型超新星, 还有几个被认为是 Ic 型的, 这并不与使用 Ia 型模板相悖, 因为本文的筛选是根据特殊性进行的离群筛选, 所以当然也会筛选出具有其他特殊性的光谱, 这样的光谱发现是偶然性的, 但大多数 Ia 型超新星候选光谱还是能够被筛选出来的.

3 结论

本文提出了一种在海量光谱中搜寻超新星候选光谱的方法, 该方法可细分为五个步骤, 第一对所有光谱进行去噪、退红移、归一化等预处理; 第二进行星系成分扣除和基于离群搜索的快速数据约减; 第

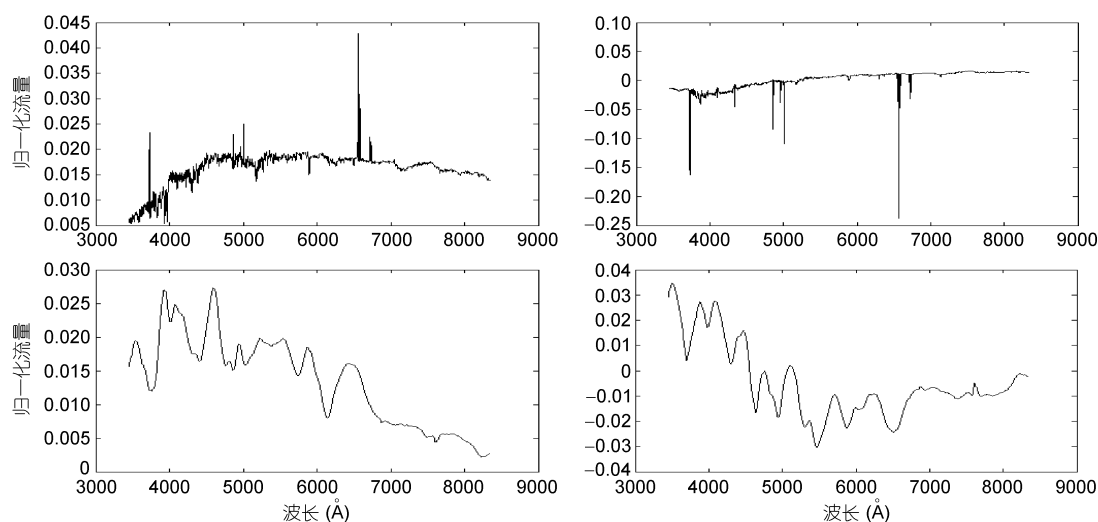
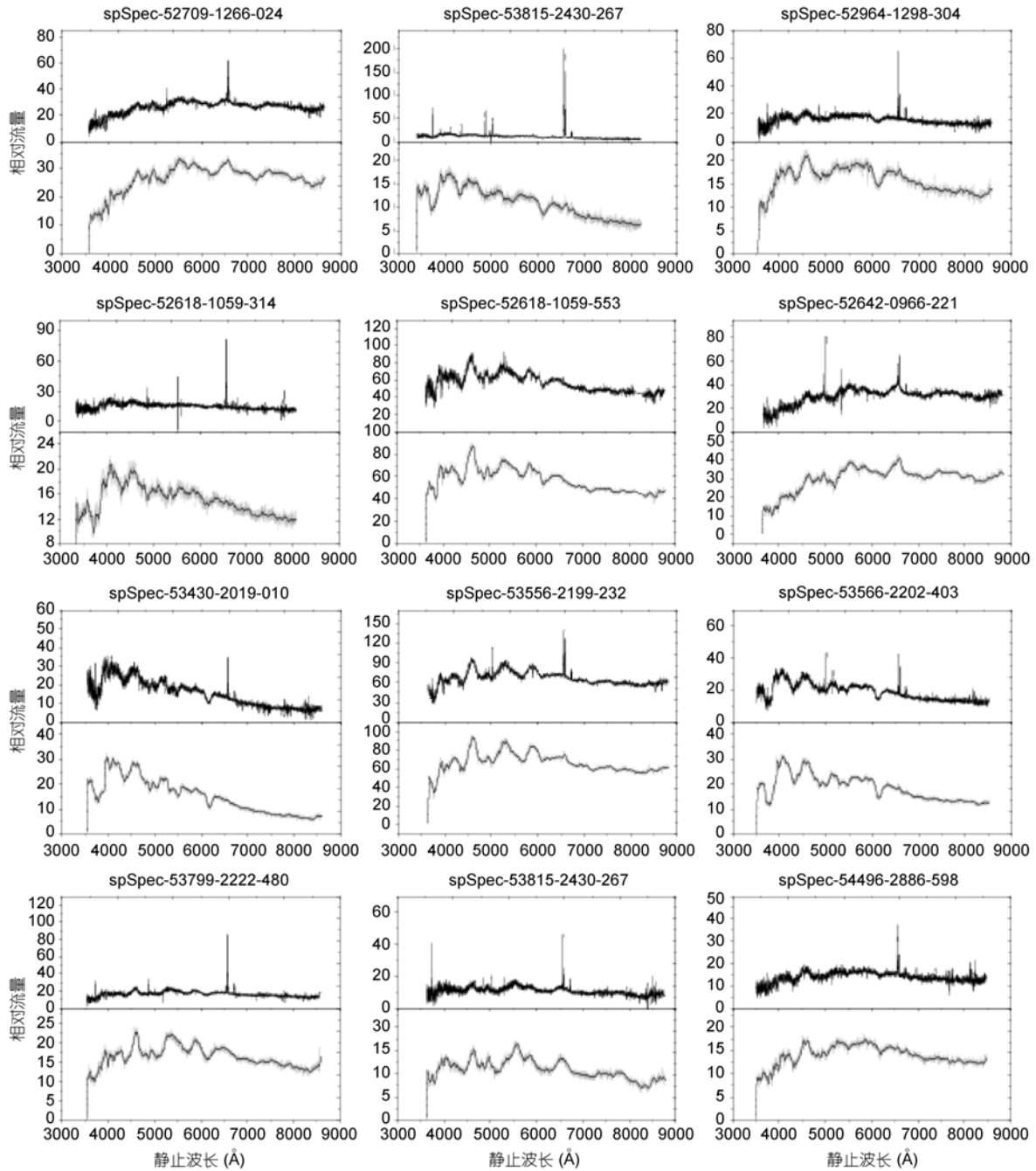


图 4 星系(上)和超新星(下)的前两个 PCA 特征谱

三通过交叉相关模板匹配方法给出匹配结果参数; 第四根据相应匹配结果的参数进行进一步筛选; 第五通过对光谱特征的人眼观察确定最终超新星候选光谱. 在将来的实时超新星搜寻中, 最后还有一个步骤是通过测光望远镜的跟踪观测进行证认. 对 SDSS-DR7 大量光谱数据的实验结果表明本文利用光谱来搜索超新星候选的方法是行之有效的, 而且所有的

优化处理也使得这一方法将来可以有效地应用到 LAMOST 的实时数据处理中, 及时获得正在爆发的超新星候选并跟踪观测证认. 不过, 特别需要指出的是, 本文方法不能保证完备性, 也就是说在具体搜索中可能会有遗漏, 原因主要有三个方面: (i) 完备性和约减比例矛盾, 因为要保证实时性的搜索, 那么约减比例就要尽可能的小, 这样会导致一些特征不明



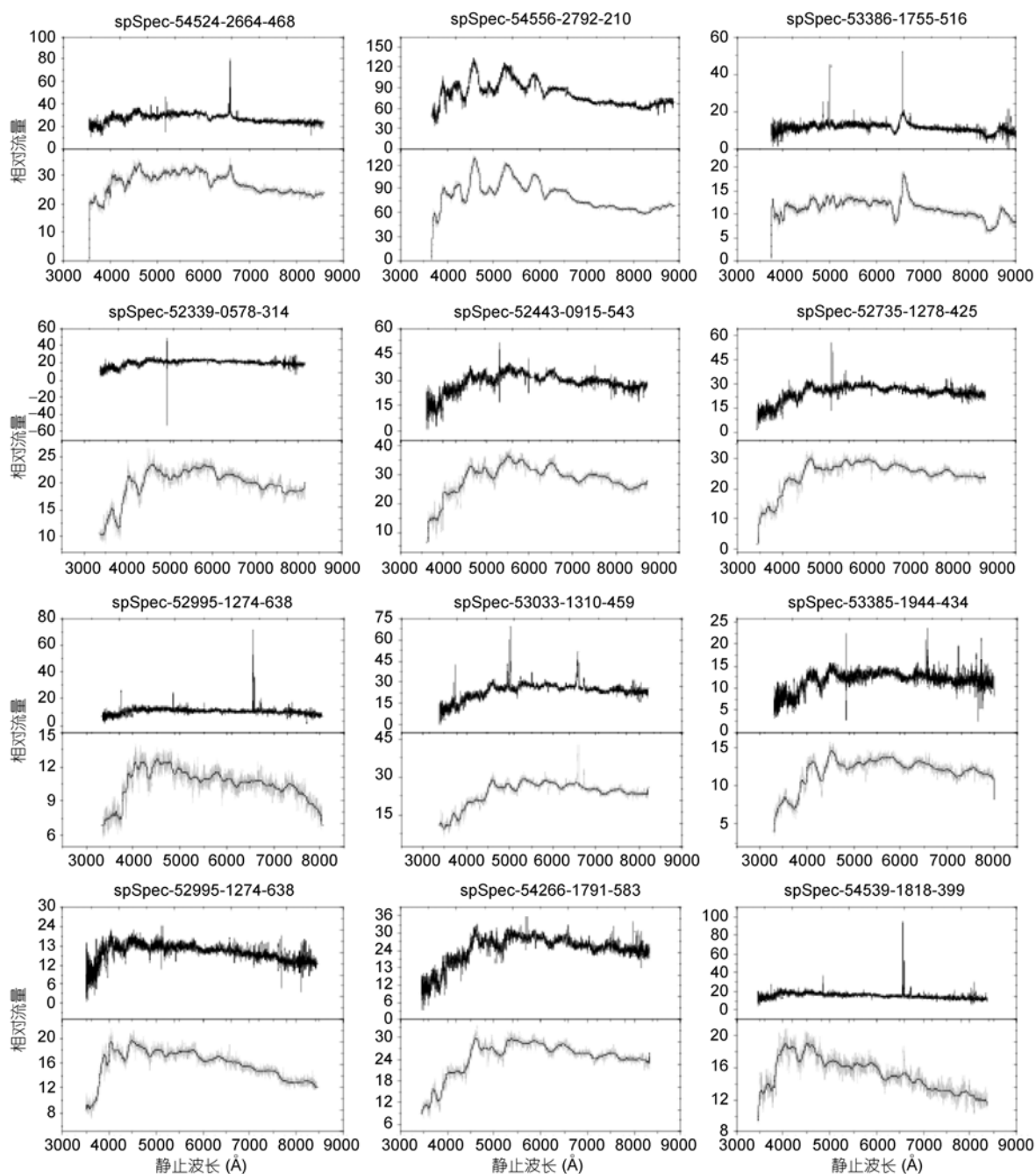


图5 本文获得的部分超新星候选体所在星系光谱(未公开记录的 24 个)

图中名称为 SDSS 光谱文件名. 每个子图中上图为 SDSS 原始光谱, 下图为去除强线后光谱(浅灰)和平滑结果(黑线)

显的超新星光谱被排除; (ii) 在匹配的过程中, 超新星的模板仍然较少, 而且类型多样, 有些差异很大, 这有可能会丢掉一些特别的超新星; (iii) 一些低信噪比的超新星光谱也可能被在某个阶段排除掉. 如何提高本文方法在低信噪比光谱中的应用也是作者在

将来需要研究的问题. 另外, 在将来的实际应用中, 对每一个超新星候选给出合理的置信度是必要的, 因为辅助的望远镜不多, 必须首先跟踪置信度较高的超新星候选目标, 这也是作者将要开展的工作. 最后, 虽然本文方法相比于光学巡天计划能够节约资

表 1 本文搜寻获得的 SDSS-DR7 中的超新星候选光谱^{a)}

SDSS 光谱文件名	光谱对应超新星名称或其母星系名称	光谱观测日期	红移	年龄	类型	r 星等
spSpec-51883-0271-171.fit	Sn 2000fx	2000-12-01	0.065	16-18	Ia-pec	17.9825
spSpec-51884-0438-462.fit	Sn2000fy	2000-12-06	0.1171	2.5~3.5	Ia-norm	17.9270
spSpec-51893-0424-355.fit	Sn2000fz	2000-12-15	0.0539	6.3~6.8	Ia-norm	16.7255
spSpec-51911-0452-319.fit	Sn2001kj	2001-01-02	0.0626	10~11	Ia-norm	17.7943
spSpec-51989-0480-024.fit	Sn2001kp	2001-03-21	0.0634	-3~-2	Ia-pec	17.5788
spSpec-51994-0500-100.fit	Sn2001kr	2001-03-26	0.0859±0.007	9.4	Ia-csm	18.1460
spSpec-52993-1304-552.fit	Sn2003ly	2003-12-17	0.0945±0.004	17.00	Ia-norm	18.4595
spSpec-53055-1744-210.fit	Sn2004ar	2004-02-20	0.0644±0.003	9.80	Ia-norm	18.0371
spSpec-53112-1462-638.fit	Sn2004co	2004-04-17	0.0293±0.005	8.1	Ia-norm	15.8176
spSpec-53149-1581-470.fit	Sn2004cp	2004-05-24	0.0534±0.005	28.5	Ia-norm	17.4909
spSpec-52618-1059-144.fit ^{N1}	anon	2002-12-10	0.0612±0.005	41.6	Ia-91bg	17.4517
spSpec-52618-1059-553.fit ^{N1}	anon	2002-12-10	0.1311±0.004	-5	Ia-norm	18.4190
spSpec-52642-0966-221.fit ^{N1}	anon	2003-01-03	0.0851±0.008	-3.4	Ia-norm	18.0071
spSpec-52709-1266-024.fit ^{N1}	anon	2003-03-11	0.1579±0.005	-8.6	Ia-norm	18.1085
spSpec-52822-1392-147.fit ^{N1}	anon	2003-07-02	0.0706±0.006	-2	Ic-norm	16.6367
spSpec-52964-1298-304.fit ^{N1}	2MASX J08390967+0724320	2003-11-21	0.0422±0.006	23.7	Ia-91T	17.2893
spSpec-53386-1755-516.fit ^{N1}	anon	2005-01-16	0.0219±0.003	34.1	IIP	18.3636
spSpec-53430-2019-010.fit ^{N1}	anon	2005-02-28	0.0767±0.003	4.3	Ia-norm	18.0901
spSpec-53556-2199-232.fit ^{N1}	MCG+03-41-035	2005-07-05	0.0398±0.005	16.1	Ia-norm	16.4611
spSpec-53566-2202-403.fit ^{N1}	anon	2005-07-07	0.0968±0.008	0.30	Ia-91T	17.8473
spSpec-53799-2222-480.fit ^{N1}	anon	2006-03-03	0.0749±0.004	19.1	Ia-norm	17.9424
spSpec-53815-2430-267.fit ^{N1}	NGC 2682	2006-03-21	0.0527±0.004	28.5	Ia-norm	18.3881
spSpec-54498-2886-598.fit ^{N1}	anon	2008-02-02	0.0859±0.007	9.4	Ia-csm	18.1183
spSpec-54524-2664-468.fit ^{N1}	NGP9 F380-0370514	2008-02-28	0.0877±0.01	0.3	Ia-norm	17.4327
spSpec-54556-2792-210.fit ^{N1}	2MASX J15053172+1759051	2008-03-31	0.0325±0.005	14.1	Ia-norm	16.2254
spSpec-51818-0358-181.fit	SN2000gb	2000-10-01	0.123	16~20	Ia	18.163952
spSpec-53033-1598-380.fit	SN2004cj	2004-01-20	0.102	-6.7	Ia-norm	17.887659
spSpec-52339-0578-314.fit ^{N2}	anon	2002-03-06	0.1372±0.0113	0	Ic-norm	17.77206
spSpec-52443-0915-543.fit ^{N2}	2dFGRS GN272Z096	2002-06-18	0.0472±0.0062	14.2	Ia-pec	17.337349
spSpec-52735-1278-425.fit ^{N2}	anon	2003-04-06	0.106		Ia	17.505713
spSpec-52995-1274-638.fit ^{N2}	anon	2003-12-22	0.1517±0.0114	-6.4	Ia-norm	18.5145
spSpec-53033-1310-459.fit ^{N2}	anon	2004-01-29	0.1219±0.0082	42.30	Ia-csm	17.531425
spSpec-53385-1944-434.fit ^{N2}	anon	2005-01-14	0.1650±0.0067	-6.7	Ia-norm	18.335785
spSpec-54086-2420-142.fit ^{N2}	anon	2006-12-15	0.0753±0.0090	-13.2	Ia-norm	18.039083
spSpec-54266-1791-583.fit ^{N2}	anon	2007-06-14	0.1125±0.0096	18.4	Ic-norm	17.512375
spSpec-54539-1818-399.fit ^{N2}	anon	2008-03-14	0.1030±0.0062	9.4	Ia-norm	18.118294

a) anon 为未命名星系; 红移: 部分为 SNID 的测量红移(包含有误差估计范围的), 其他为 SDSS 红移; 年龄指 SNID 最匹配模板的年龄; 上角“^{N1}”为新发现并已报告的, 上角“^{N2}”为本文首次发现的

源, 但由于光谱信息受噪声影响很大, 预期的搜索数量估计会小于许多大型的超新星巡天计划, 将来在

具体 LAMOST 的使用过程中仍然需要进行适当的改进和调整.

致谢

在此感谢国家天文台胡景耀研究员, 袁予雷研究员, 邓劲松博士和吴朝博士, 通过与他们有意义的讨论和他们的建议增加了作者对天文工作的理解, 另外感谢 Peter Nugent 先生公开提供其超新星理论模板.

参考文献

- 1 汲培文, 胡景耀. 超新星的观测研究及其科学意义——兼论基础研究选题的指导思想. 中国科学基金, 1997, 11(3): 175—178
- 2 Frieman J A, Bassett B, Becker A, et al. The sloan digital sky survey-II supernova survey: Technical summary. *Astron J*, 2008, 135(1): 338—347
- 3 Sako M, Bassett B, Becker A, et al. The sloan digital sky survey-II supernova survey: Search algorithm and follow-Up Observations. *Astron J*, 2008, 135(1): 348—373
- 4 Kuznetsova N, Barbary K, Connolly B, et al. A new determination of the high-redshift type Ia supernova rates with the hubble space telescope advanced camera for surveys. *Astrophys J*, 2008, 673(2): 981—998
- 5 Filippenko A V, Li W D, Treffers R R, et al. The lick observatory supernova search with the katzman automatic imaging telescope. In: *Astron. Soc. Pacific, Conf. Ser. 246: IAU Colloq. 183: Small Telescope Astronomy on Global Scales*. San Francisco: Astronomical Society of the Pacific, 2001.121—130
- 6 Hamuy M, Folatelli G, Nidia I, et al. The carnegie supernova project: The low-redshift survey. *Publ Astron Soc Pacific*, 2006, 118(839): 2—20
- 7 Aldering G, Adam G, Antilogus P, et al. Overview of the nearby supernova factory. In: *Proceedings of the SPIE, USA*. Berkeley: Lawrence Berkeley National Laboratory, 2002, 4836: 61—72
- 8 Kaiser N, Aussel H, Burke B E, et al. Pan-STARRS: A large synoptic survey telescope array. In: *Proceedings of the SPIE, USA*. Berkeley: Lawrence Berkeley National Laboratory, 2002, 4836: 154—164
- 9 Blondin S, Tonry J L. Determining the type, redshift, and phase of a supernova spectrum. In: *Proceedings of AIP Conference, Italy*, 2007, 924: 312—321. Doi: 10.1063/1.2774875
- 10 Blondin S, Tonry J L. Determining the type, redshift, and age of a supernova spectrum. *Astrophys J*, 2007, 666(2): 1024—1047
- 11 Tu L P, Luo A L, Wu F C, et al. New supernova candidates from the SDSS-DR7 spectral survey. *Res Astron Astrophys*, 2009, 9(6): 635—640
- 12 李卫东, 赵昭旺, 裘予雷, 等. 北京天文台超新星巡天系统——I. 巡天设备及软件. 中国科学 A 辑: 数学 物理学 天文学, 1998, 41(3): 283—288
- 13 李卫东, 裘予雷, 乔琪源, 等. 北京天文台超新星巡天系统——II. 1996 年巡天结果. 中国科学 A 辑: 数学 物理学 天文学, 1998, 41(5): 464—470
- 14 屠良平, 罗阿理, 吴福朝, 等. 海量星系光谱中的超新星候选范围自动约减. 光谱学与光谱分析. 2009, 29(12): 3420—3423
- 15 Donoho D L. De-noising by soft-thresholding. *IEEE Trans Inf Theory*, 1995, 41(3): 613—627
- 16 Duan F Q, Liu R, Guo P, et al. Automated spectral classification using template matching. *Res Astron Astrophys*, 2009, 9(3): 341—348
- 17 Madgwick D S, Hewett P C, Mortlock D J, et al. Spectroscopic detection of type Ia supernovae in the sloan digital sky survey. *Astrophys J*, 2003, 599: L33—L36
- 18 Richard O D, Peter E H, David G S. *Pattern Classification*. 2ed. New York: Wiley-Interscience, 2000
- 19 Abazajian K, Adelman-McCarchy J K, Agueros M A, et al. The first data release of the sloan digital sky survey. *Astron J*, 2003, 126(4): 2081—2086
- 20 Howell D A, Wang L. Automated supernova classification. *Bull Am Astron Soc*, 2002, 34: 1256
- 21 Howell D A, Sullivan M, Perrett K, et al. Gemini spectroscopy of supernovae from the supernova legacy survey: Improving high-redshift supernova selection and classification. *Astrophys J*, 2005, 634(2): 1190—1201
- 22 Lidman C, Howell D A, Folatelli G, et al. Spectroscopic confirmation of high-redshift supernovae with the ESO VLT. *Astron Astrophys*, 2005, 430: 843—851
- 23 Hook I M, Howell D A, Aldering G, et al. Spectra of high-redshift type Ia supernovae and a comparison with their low-redshift counterparts. *Astron J*, 2005, 130(6): 2788—2803
- 24 Zheng C, Romani R W, Sako M, et al. First-year spectroscopy for the sloan digital sky survey-II supernova survey. *Astron J*, 2008, 135(5): 1766—1784
- 25 Breunig M M, Kriegel H P, Ng R T, et al. LOF: Identifying Density-Based Local Outliers. *ACM SIGMOD Record*, 2000, 29: 93—104
- 26 Tonry J, Davis M. A survey of galaxy redshifts. I. Data reduction techniques. *Astron J*, 1979, 84(10): 1511—1525
- 27 Abazajian K N, Adelman-McCarchy J K, Agueros M A, et al. The seventh data release of the sloan digital sky survey. *Astrophys J Suppl Ser*, 2009, 182(2): 543—558
- 28 Yip C W, Connolly A J, Szalay A S, et al. Distributions of galaxy spectral types in the sloan digital sky survey. *Astron J*, 2004, 128(2): 585—609
- 29 Nugent P. *The Non-LTE Spectrum Synthesis of Type Ia Supernovae*. Dissertation for the Doctoral Degree. Norman: University of Oklahoma, 1997