

基于 Logistic 回归模型的 Blazar 天体的分类*

唐洁¹⁾ 吴学兵^{2)†}

1) (陕西理工学院物理系, 汉中 723001)

2) (北京大学物理学院天文系, 北京 100871)

(2011 年 3 月 24 日收到; 2011 年 5 月 16 日收到修改稿)

从文献中收集了 205 个 Blazar 天体, 包括 142 个 BL Lac 天体和 63 个平谱射电类星体 (FSRQs). 对这些天体的类别与它们的红移、射电 5 GHz 辐射流量、光学 V 波段流量、1 keV 处 X 射线流量、X 射线光子谱指数进行了相关性和 Logistic 回归分析. 结果表明, 对 Blazar 天体分类产生主要影响的因素是红移、射电 5 GHz 辐射流量和 X 射线光子谱指数, 综合应用这三个物理量判别 Blazar 天体的分类的准确率可达到 91.2%, 得到的分类方程具有良好的预测效果, 可以作为 Blazar 天体分类的一个重要的判据. 而光学 V 波段流量和 1 keV 处 X 射线流量不能区分 BL Lac 天体和 FSRQs, 它们与 Blazar 天体分类没有相关性. 本文结果支持将 BL Lac 天体和 FSRQs 归为 Blazar 天体, 不同类别的 Blazar 天体之间能通过一种演化序列相联系.

关键词: Logistic, 分类, Blazar 天体

PACS: 98.54.Cm, 98.62.Ve, 02.50.-r, 98.52.Cf

1. 引言

Blazar 天体是活动星系核 (AGNs) 中性质较为独特的一个子类, 具有大幅快速光变、高的光学偏振、视超光速运动和非热连续辐射等特性^[1]. Blazar 天体分为 BL Lac 天体和平谱射电类星体 (FSRQs), 随着观测和研究的深入, 分类越来越细, BL Lac 天体和 FSRQs 又分为许多亚类, 亚类之间又存在着交叠和融合. 这些分类可能只是表象, 也许隐含着统一的物理过程. 因此 Blazar 天体的分类与统一还存在许多问题, 有待进一步研究.

由于 BL Lac 天体和 FSRQs 的辐射都是非热的并具有相对论聚束效应, 人们把它们归为 Blazar 天体^[2]. 但从文献上的统计数据看, BL Lac 天体和 FSRQs 又各自具有不同特点, 针对这个问题不同的学者从不同角度进行了探讨. Ghisellini 等^[3]认为 Blazar 天体序列可以用热光度这个参数来统一. Fan 等^[4,5]对 BL Lac 天体和 FSRQs 的辐射流量密度和线偏振度变化特点进行了研究, 在光学性质上, 发现 BL Lac 天体和 FSRQs 没有明显的区别. Yang 等^[6]讨论了 148 个 Blazar 天体的发射线光度、偏振

度、核主导系数之间的关系, 认为 BL Lac 天体和 FSRQs 应归为同一类型. Xie 等^[7]分析了 BL Lac 天体和 FSRQs 的星等和红移的关系, 红移和发射线分布以及软 X 射线谱斜率, 得到 BL Lac 天体和 FSRQs 不应归为 Blazar 天体的结论. Sambruna 等^[8]对 Blazar 天体多波段的能谱进行了讨论, 得到的结论是 BL Lac 天体和 FSRQs 的宽波段谱指数有着显著的差异. Zhang 等^[9]分析了 29 个有 γ 噪声的 Blazar 天体, 发现 BL Lac 天体和 FSRQs 的 γ 辐射机制可能不同. 这些研究成果由于侧重点不同, 对 BL Lac 天体和 FSRQs 的物理属性的异同提出了两种不同看法: 一种认为 BL Lac 天体和 FSRQs 之间存在区别, 不能归为一类; 另一种却认为应该将它们统一为 Blazar 天体. 实际上影响 Blazar 天体分类的因素是多方面的, 如果综合考虑这些因素或许能更好地解释不同类别的 Blazar 天体之间的联系和区别.

现有的分类方法很多, West^[10]将 5 种神经网络模型和 5 种统计分类模型进行了比较, 结果表明, Logistic 回归模型在这 10 种方法中分类准确率最高. 本文结合其他文献的研究成果, 以影响 Blazar 天体分类的因素为基础选取变量, 利用二元 Logistic 回归模型从 Blazar 天体的红移、射电 5 GHz 辐射流量、

* 国家自然科学基金重点项目 (批准号: 11033001) 和国家重点基础研究发展计划 (批准号: 2007CB815405) 资助的课题.

† 通讯联系人. E-mail: wuxb@bac.pku.edu.cn

光学 V 波段流量、1 keV 处 X 射线流量、X 射线光子谱指数等因素探寻 BL Lac 天体和 FSRQs 内在的联系,找出影响它们分类的主要因素,并根据建立的 Logistic 回归模型预测某一未知分类的 Blazar 天体属 BL Lac 天体还是 FSRQs,再验证不同类型的 Blazar 天体是否能够通过一种演化序列相联系.

2. Logistic 回归模型

Logistic 回归方法属于非线性回归方法,它研究某一现象发生的概率以及概率大小取决于哪些因素的方法^[11].对于因变量要求是离散的分类变量,自变量可以是连续的,也可以是离散的.

典型的 Logistic 函数的形式是^[11]

$$p_i = \frac{1}{1 + e^{-y_i}} = \frac{1}{1 + e^{-(\alpha + \beta x_i)}}, \quad (1)$$

其中 x_i, y_i 分别代表自变量,因变量, y_i 取 0 或 1, α 和 β 分别表示回归截距和回归系数, p_i 为某现象发生的概率.

对(1)式再做变换,得到 Logistic 回归模型^[11]

$$\ln\left(\frac{p_i}{1 - p_i}\right) = y_i = \alpha + \beta x_i. \quad (2)$$

由(2)式可知 $\ln\left(\frac{p_i}{1 - p_i}\right)$ 是 x_i 的线性函数, $\frac{p_i}{1 - p_i}$ 表示现象发生与不发生的概率比, α, β 用极大似然参数估计求得,进而推知现象发生与不发生的概率.

Logistic 回归模型巧妙地将预测某现象在 $[0, 1]$ 范围内发生概率的问题转化为预测某现象发生的概率比问题,因而这种方法比较准确^[11].

3. Blazar 天体的分类

3.1. 研究样本的选择

Donato 等^[12]给出了一个包含有 268 个源的

Blazar 样本,并提供了部分源的红移、射电 5 GHz 辐射流量(以下简称射电辐射流量)、光学 V 波段流量、1 keV 处 X 射线流量(以下简称 X 射线流量)、X 射线光子谱指数(以下简称 X 射线谱指数).为了研究问题方便,我们从这些样本中选取一个新样本进行研究.由于进行 Logistic 回归分析时自变量里有红移,所以去掉样本中所有缺红移数据的 57 个源,同时除去谱指数或流量的标准化残差不在 $[-2, 2]$ 区间的 6 个源,以免这些有特异值的源影响到模型的建立^[11].对于有多个谱指数的源,选择第 1 个谱指数来分析. Donato 等^[12]的样本里将 BL Lac 天体分为高峰频 BL Lac 天体(HBLs)和低峰频 BL Lac 天体(LBLs),本文为了研究方便将它们归为 BL Lac 天体一类,最后得到了一个有 205 个 Blazar 源的有效样本,样本中含 142 个 BL Lac 天体(96 个 HBLs, 46 个 LBLs),63 个 FSRQs.

3.2. 相关性和多重共线性

选择 Blazar 天体的两个子类 BL Lac 天体和 FSRQs 为因变量,以红移、射电辐射流量、光学 V 波段流量、X 射线流量、X 射线谱指数为自变量进行 Logistic 回归分析.这些自变量间可能存在完全或近似完全的线性关系,也就是多重共线性.要使 Logistic 回归分析得到的回归系数和显著性概率有意义,必须确保自变量之间没有强的多重共线性^[11].

将相关系数、方差膨胀因子(VIF)、容忍值作为多重共线性诊断统计量.从表 1 相关系数的计算结果可以看出,除 V 波段流量和 X 射线谱指数之间的相关系数为 0.667,相关性较强,其余的均小于 0.5;五个变量的容忍值都远大于 0 接近 1,越接近 1 说明自变量间的共线性越弱;而方差膨胀因子 VIF 没有超过 3,都远小于 10,共线性不是很明显,不需要进行多重共线性的处理.

表 1 相关系数和多重共线性统计

	相关系数					共线性统计量	
	红移	射电辐射流量	V 波段流量	X 射线流量	X 射线谱指数	容忍值	VIF
Blazars 类别	-0.624	-0.395	0.039	0.115	0.449		
红移	1.000	0.073	-0.210	-0.168	-0.324	0.855	1.169
射电流量	0.073	1.000	0.452	0.126	-0.216	0.664	1.506
V 波段流量	-0.210	0.452	1.000	0.667	0.122	0.389	2.569
X 射线流量	-0.168	0.126	0.667	1.000	0.114	0.516	1.939
X 射线谱指数	-0.324	-0.216	0.122	0.114	1.000	0.828	1.207

3.3. 数据分析

为了让分析结论更加可靠,需要建立不同的 Logistic 回归方程,比较它们的最终预测结果,依据分类准确性选择一个最佳模型.

将红移、射电辐射流量、V 波段流量、X 射线流量、X 射线谱指数 5 个自变量分别进行单变量二元 Logistic 回归分析.在分析前需要选择一个分类概率临界值,比如选 0.5,规定凡是分类概率值大于 0.5 的源是 BL Lac 天体,小于 0.5 的源是 FSRQs.图 1

分别是五个物理量与概率的分类图,通过分类图很容易找到误判率最低的分类概率临界值.图中的虚线是概率分割线,它们的临界值分别为 0.67,0.55,0.69,0.67,0.55,V 波段流量和 X 射线流量的显著性水平 P 值分别为 0.581,0.136,其余三个自变量的 P 值均小于 10^{-6} ,一般要求 P 值小于 0.05^[11],引入变量后分类的准确率分别为:83.9%,79%,48.3%,54.6%,81.5%.从分类准确率和显著性水平来看,V 波段流量和 X 射线流量与 Blazar 天体分类的关联性不大,其他的都有统计意义.

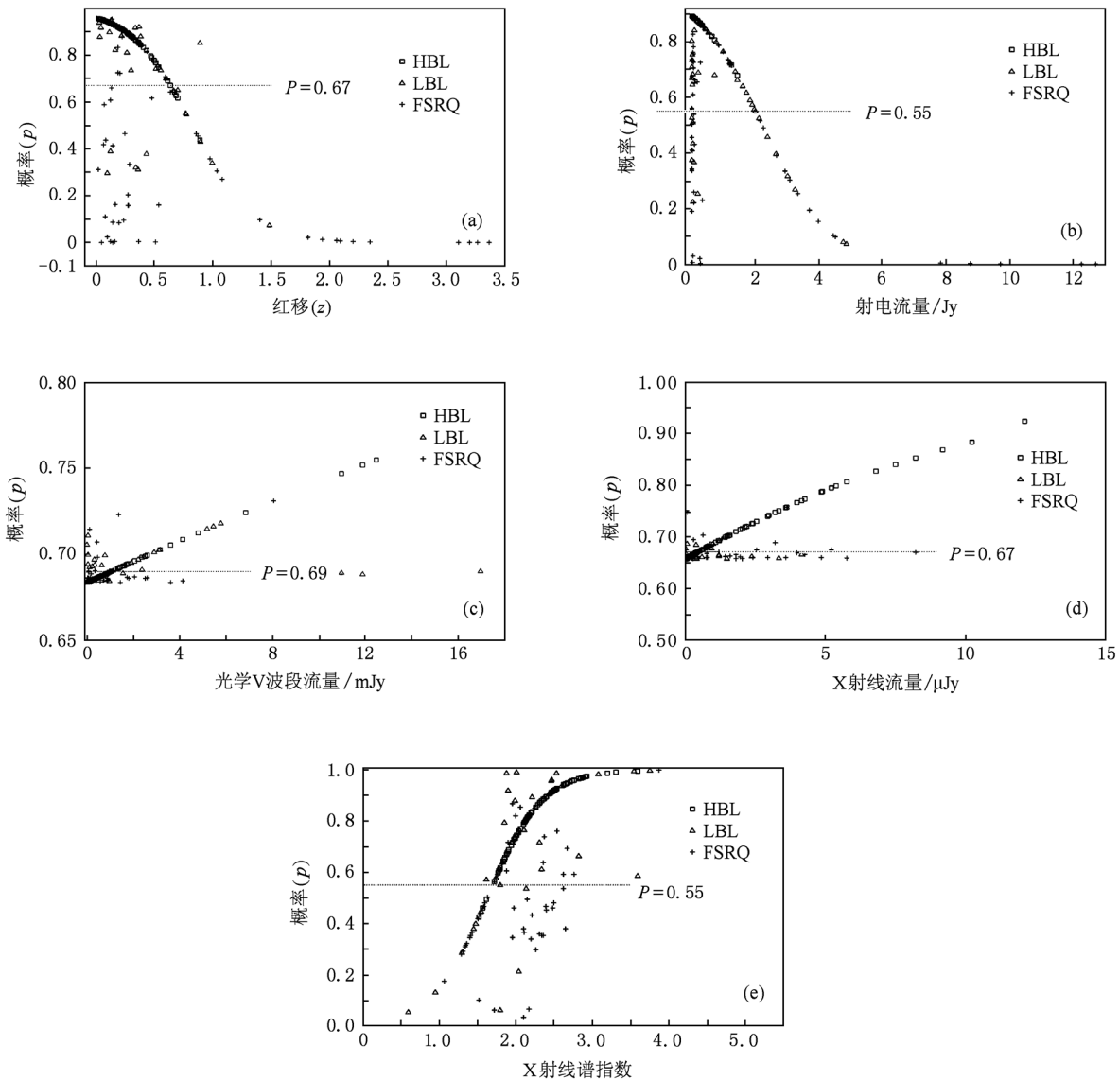


图 1 分类图 (a) 红移;(b) 射电辐射流量;(c) 光学 V 波段流量;(d) X 射线流量;(e) X 射线谱指数

为了考察所有自变量对因变量的分类影响,将所有自变量强制进入分类方程进行多变量 Logistic 回归分析.V 波段流量和 X 射线流量的显著性水平

P 值分别为 0.153,0.653,都远大于显著性水平 0.05,这和单变量分析结果是一致的,它们本应该不能进入方程,但在这里由于用的是不筛选自变量的

策略,这些变量也被强制进入方程,其他三个变量和常量的 P 值都远小于 0.05. 当分类概率临界值选为 0.55 时,其分类准确率为 91.2%, BL Lac 天体和 FRSQ 分别只有 9 个判错,准确率还是很高的,分类方程为:

$$p = \frac{1}{1 + e^{(-4.395x_1 - 0.72x_2 - 0.212x_3 + 0.071x_4 + 1.805x_5 + 1.743)}}, \quad (3)$$

$x_1, x_2, x_3, x_4, x_5, p$ 分别为红移、射电辐射流量、V 波段流量、X 射线流量、X 射线谱指数、概率. 由于方程包含不显著性解释变量,因此该模型不可用.

Logistic 回归模型可以通过比较预测结果和实际结果的符合程度,判别分类方程的正确性. 为了使模型更具合理性,用来判断模型好坏与拟合模型的数据不能是同一批数据. 因此选择 175 个 Blazar 源的数据作训练样本来建立回归模型,并用得到的分类方程预测剩下的 30 个源作检验样本的分类,为避免选择效应,这 30 个源选择 Donato 等^[12] 文献列表里前 20 个 BL Lac 天体和前 10 个 FSRQs.

由于自变量包含了不显著性解释变量,在新建模过程中,解释变量的筛选采用基于极大似然估计的逐步筛选策略. 多变量逐步 Logistic 回归分析结果只有红移、射电辐射流量、X 射线谱指数进入模型, V 波段流量和 X 射线流量两个变量没有引入方程,从单变量 Logistic 回归分析也可知 V 波段流量和 X 射线流量对分类准确率特别低. 除去这两个变量,选择分类概率临界值为 0.65, 设 $x_1 =$ 红移, $x_2 =$ 射电辐射流量, $x_3 =$ X 射线谱指数,得到 Blazar 天体的分类的 Logistic 回归概率方程:

$$p = \frac{1}{1 + e^{(-4.012x_1 - 0.677x_2 + 2.394x_3)}}. \quad (4)$$

训练样本里含有 122 个 BL Lac 天体, 53 个 FSRQs, 由该方程得到分类为 BL Lac 天体的为 113 个源,准确率为 92.6%, 分类为 FSRQs 的为 46 源,准确率为 86.8%, 共只有 16 源判错,总的准确率为 90.9%. 将测试样本的 30 个源相关数据代入回归概率方程,如表 2 所示,测试得到的结论只有 2 例判错

表 2 Blazar 天体测试样本

IAU 名	z	F_R/Jy	F_O/mJy	$F_X/\mu\text{Jy}$	Γ	概率值	预测分类	实际分类
0014 + 813	3.366	0.551	0.91	0.35	1.89	0.000	F	F
0016 + 731	1.781	1.750	0.06	0.05	1.43	0.007	F	F
0038 - 020	1.178	0.581	0.03	0.10	1.40	0.146	F	F
0048 - 097	0.200	2.743	1.75	0.42	1.73	0.815	B	B
0110 + 418	0.096	0.036	0.39	0.30	2.11	0.990	B	B
0112 - 017	1.365	1.200	0.23	0.15	1.57	0.074	F	F
0118 - 272	0.557	1.284	1.56	0.28	2.74	0.969	B	B
0120 + 340	0.272	0.034	3.60	2.42	1.96	0.973	B	B
0122 + 0903	0.339	0.001	0.05	0.03	1.62	0.925	B	B
0133 + 476	0.859	2.920	0.47	0.30	1.92	0.304	F	F
0145 + 138	0.125	0.005	0.26	0.22	2.32	0.994	B	B
0152 + 017	0.080	0.058	0.90	0.47	2.48	0.996	B	B
0158 + 003	0.299	0.011	0.23	1.21	2.46	0.991	B	B
0202 + 149	0.405	2.400	0.02	0.06	0.68	0.165	F	F
0205 + 3509	0.318	0.004	0.18	0.15	2.71	0.995	B	B
0208 - 512	1.003	3.311	0.59	0.61	2.04	0.201	F	F
0212 + 735	2.370	2.198	0.45	0.26	0.66	0.000	F	F
0214 + 517	0.049	0.271	1.55	1.60	2.04	0.989	B	B
0219 + 428	0.444	1.040	4.55	0.72	2.39	0.962	B	B
0235 + 164	0.940	3.336	2.58	0.15	1.75	0.137	F	B**
0237 - 233	2.223	3.520	0.81	0.31	1.62	0.001	F	F
0257 + 3429	0.245	0.010	0.14	0.02	2.67	0.996	B	B
0317 + 1834	0.190	0.017	0.21	0.17	2.32	0.992	B	B
0323 + 022	0.147	0.042	0.90	6.81	2.16	0.990	B	B
0331 - 365	0.308	0.009	0.24	0.34	2.20	0.982	B	B
0332 - 403	1.445	2.600	0.14	0.14	1.60	0.024	F	F
0347 - 121	0.188	0.009	0.19	2.05	2.17	0.988	B	B
0350 - 3712	0.165	0.017	0.59	0.35	2.17	0.989	B	B
0406 + 121	1.020	0.746	0.12	0.05	1.79	0.423	F	B**
0414 + 009	0.287	0.070	0.85	5.22	2.54	0.992	B	B

(B 代表 BL Lac 天体, F 代表 FSRQs, 上标的双星号表示判错的源), 准确率高达 93.3%. 表 2 列出了测试源的数据, 表的第一列为 IAU 名, 第二列为红移 z , 第三列为射电辐射流量 F_R , 第四列为光学 V 波段流量 F_o , 第五列为 X 射线流量 F_x , 第六列为 X 射线谱指数 Γ , 第七列为分类为 BL Lac 天体的概率值, 第八列为预测分类, 第九列为实际分类. 用该回归模型对于整个 205 个样本只有 18 例判错, 准确率为 91.2%, 验证了回归模型方程用来判别 Blazar 天体分类是可行的.

4. 讨 论

从变量间的相关系数来看, Blazar 天体分类与 V 波段流量、X 射线流量的相关系数分别为 0.039, 0.115, 都偏小, 在一定程度上说明 BL Lac 天体和 FSRQs 的 V 波段流量、X 射线流量不存在较明显的区别. 在单变量 Logistic 回归分析中分别引入这两个量时, 分类的准确率仅为 48.3% 和 54.6%; 在多变量 Logistic 回归分析中由于它们的显著性水平 P 值不满足要求, 导致得到的回归方程没有意义; 在多变量逐步 Logistic 回归分析过程中将它们排除在最终的 Logistic 回归方程外, 五个物理量全部进入和除去这两个物理量的预测方程的准确率是一样的, 说明这两个物理量对分类没有贡献. 这些结果都相互印证了 Blazar 天体的分类与 V 波段流量、X 射线流量关系不大. 这个结论和 Fan 等^[4,5] 认为不同类型的 Blazar 天体的辐射流量密度没有明显区别是一致的. 此外, 由于辐射流量的变化与光变时标存在联系^[1], 不同类型的 Blazar 天体具有周期性光变的共同特点也能得到合理解释^[13]. 用 V 波段流量、X 射线流量不能区分开 BL Lac 天体和 FSRQs, 说明它们没有明显的区别, 可以用这两个参数来统一它们为 Blazar 天体.

单变量 Logistic 回归分析结果表明红移、射电辐射流量、光学 V 波段流量、X 射线流量、X 射线谱指数这五个变量, 在控制了其他变量组影响之后, 红移、射电辐射流量、X 射线光子谱指数对 Blazar 天体的分类具有单独的影响作用, 这些结果也可从图 1 得到. 在图 1(a), (b), (e) 中, 分割线把图分成两个区域, 位于分割线上的区域主要是 BL Lac 天体, 下面的区域主要是 FSRQs; 图 1(c), (d) 中这些天体重叠在一起, 很难区分开它们. 从图 1(a), (b) 可见

随红移、射电辐射流量增大, Blazar 天体分为 BL Lac 天体的概率反而减小, 图 1(e) 恰好相反, 随 X 射线谱指数增大, 样本源分为 BL Lac 天体的概率增大. 从多变量逐步 Logistic 回归模型的方程也印证了这一点, 概率回归方程中红移、射电辐射流量前的系数是负的, BL Lac 天体的概率随它们的增大反而减小, X 射线光子谱指数前的系数是正的, 随它们的增大而增大.

多变量逐步 Logistic 回归分析将三组变量联合起来预测分类的准确率达到 91.2%, 高于五个变量单独预测的准确率, 具体情况见图 2 所示. 图中 BL Lac 天体和 FSRQs 的数目指用回归方程判对的源的个数. 由柱状图对比可知对 Blazar 天体分为 BL Lac 天体和 FSRQs 两类起关键作用的因素主要集中在红移、射电辐射流量、X 射线光子谱指数上, 特别是红移和 X 射线光子谱指数, 仅依靠这两个物理量区分 BL Lac 天体和 FSRQs 的准确率高达 83.9% 和 81.5%. 进一步证实了 Xie 等^[7] 的研究结论: BL Lac 天体和 FSRQs 有不同的软 X 射线谱斜率, 不同的红移和发射线分布; 也支持 Cao 等^[14] 的看法. Donato 等^[12] 也研究了这五个物理量各自的分布特性, 从他们的分析结果来看, BL Lac 天体和 FSRQs 的 V 波段流量、X 射线流量的分布规律没有明显区别, 而红移、射电辐射流量、X 射线光子谱指数三个物理量存在区别, 也和我们的结论相符合. 但他们的结果只整体给出了 BL Lac 天体和 FSRQs 存在统计区别, 没有给出某个源到底在多大程度上区别于另一个源, 以及多大程度上属于 BL Lac 天体和 FSRQs.

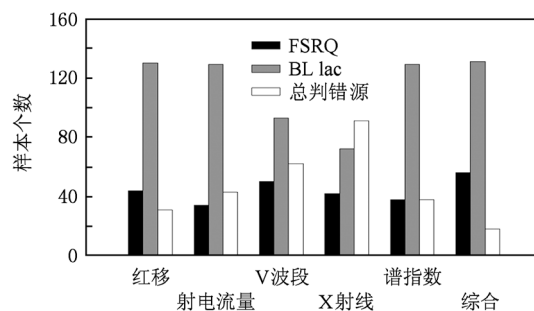


图 2 BL Lac 天体, FSRQs 和总判错源的分布图

BL Lac 天体和 FSRQs 可分为诸多子类, 根据同步辐射峰频的不同, BL Lac 天体可分为 HBLs 和 LBLs, FSRQs 包括光学激变类星体、高偏振类星体、核主导类星体. 我们选择的 Blazar 天体样本里包括

这些子类,通过 Logistic 回归模型将这些子类归为 BL Lac 天体和 FSRQs 两大类,说明光学激变类星体、高偏振类星体、核主导类星体在红移、射电辐射流量和 X 射线光子谱指数上的差别可以忽略,也证明 HBLs 和 LBLs 尽管在能谱峰频有高低区别,但可以归为同一类型的 BL Lac 天体. 从图 1(a), (b), (e) 来看, HBLs 有较小的红移和射电辐射流量、较大的 X 射线谱指数和概率值; 大部分 FSRQs 有较大的红移、较小的 X 射线谱指数和概率值; LBLs 和小部分 FSRQs 介于两者之间; HBLs 的分布与 LBLs, FSRQ 有明显差异,而后两者却较接近. LBLs 似乎是 HBLs 和 FSRQs 的过渡状态, BL Lac 天体出现分类错误的源大多数是 LBLs, 凡是红移大于 0.65, 射电辐射流量大于 2.0 Jy, X 射线谱指数小于 1.70 的 LBLs 归为 FSRQs, 其余部分归为 BL Lac 天体. 所有的 HBLs 都密集在一个很小的区域, 而 LBLs 和 FSRQs 较分散, 整个分类图出现随概率的增大, FSRQ, LBL, HBL 依序列 FSRQ-LBL-HBL 排列, 说明从 FSRQ 到 LBL 到 HBL 的演变确是连续的, 三者之间的差异有一种连续变化的趋势, 表明其中可能具有相同的物理过程, 这种趋势可能是一种演化序列, 这与 Ghisellini 等^[3] 和 Fossati 等^[15] 的结论是一致的, 更进一步确切地证实了 Sambruna^[16] 认为 LBLs 是 FSRQs 到 HBLs 的中间状态.

5. 结 论

本文基于 Logistic 回归模型建立了一种 Blazar 天体的分类模型, 从模型分类准确率和对新数据

的预测能力来看, Logistic 回归方法对于 Blazar 天体的分类具有良好的拟合和预测效果.

BL Lac 天体和 FSRQs 既存在联系, 又有明显区别^[17-20]. 我们对 Blazar 天体的类别与它们的红移、射电辐射流量、光学 V 波段流量、X 射线流量、X 射线谱指数进行了相关性和 Logistic 回归分析. 从分析结果看, BL Lac 天体和 FSRQs 的光学 V 波段流量和 X 射线流量相似, 没有明显的区别, 可以用这两个参数来统一它们为一类天体. 同时发现它们的红移、射电辐射流量和 X 射线光子谱指数相差较大, 是 Blazar 天体分为 BL Lac 天体和 FSRQs 的主要影响因素, 因此 BL Lac 天体和 FSRQs 也存在区别. 支持将 BL Lac 天体和 FSRQs 作为 Blazar 天体的两个子类的观点, 它们的内部辐射机制可能存在差异, 不同类别的 Blazar 天体能通过一种演化序列 FSRQ-LBL-HBL 相联系. 当考虑光学 V 波段流量和 X 射线流量时, 它们趋向于 Blazar 天体的统一模型, 从而也支持活动星系核的统一模型.

建立了 Blazar 天体分为 BL Lac 天体和 FSRQs 的判别模型, 应用该模型可以对某一未知分类的 Blazar 天体进行预测. 若要预测某一 Blazar 天体是 BL Lac 天体还是 FSRQs, 只要将红移、射电辐射流量、X 射线光子谱指数代入 Logistic 回归概率方程, 就可以知道该天体分类. 利用 Logistic 回归模型相当好的分类能力和预测能力, 实现了对新观测或待判样本源的分类与预测, 这在活动星系核观测中将具备较强的指导意义.

感谢审稿专家对本文提出的宝贵意见和建议.

- [1] Urry C M, Padovani P 1995 *Pub. Astron. Soc. Pacif.* **107** 803
- [2] Angel J R, Stockman H S 1980 *Ann. Rev. Astron. Astrophys.* **18** 321
- [3] Ghisellini G, Celotti A, Fossati G, Maraschi L, Comastri A 1998 *Mon. Not. R. Astron. Soc.* **301** 451
- [4] Fan J H, Yuan Y H, Liu Y 2008 *Pub. Astron. Soc. Jpn.* **60** 707
- [5] Fan J H 2003 *Astrophys. J.* **585** L23
- [6] Yang J H, Fan J H, Yang R S 2010 *Sci. China Phys. Mech. Astron.* **53** 1162
- [7] Xie G Z, Zhang Y H, Fan J H 1993 *Astron. Astrophys.* **278** 6
- [8] Sambruna R M, Maraschi L, Urry C M 1996 *Astrophys. J.* **463** 444
- [9] Zhang X, Xie G Z, Zhao G, Ma L, Yi J D, Bai J M 2001 *Acta Phys. Sin.* **50** 353 (in Chinese) [张 雄, 谢光中, 赵 刚, 马力, 伊继东, 白金明 2001 物理学报 **50** 354]
- [10] West D 2000 *Computer & Operations Research* **27** 1131
- [11] Wang J C, Guo Z G 2001 *Logistic Regression Model: Method and Application* (Beijing: Higher Education Press) p8 (in Chinese) [王济川, 郭志刚 2001 Logistic 回归模型: 方法与应用 (北京: 高等教育出版社) 第 8 页]
- [12] Donato D, Ghisellini G, Tagliaferri G, Fossati G 2001 *Astron. Astrophys.* **375** 739
- [13] Tang J, Zhang X 2010 *Acta Phys. Sin.* **59** 7516 (in Chinese) [唐 洁, 张 雄 2010 物理学报 **59** 7516]
- [14] Cao X W, Jiang D R 1999 *Mon. Not. R. Astron. Soc.* **307** 802

- [15] Fossati G, Maraschi L, Celotti A, Comastri A, Ghisellini G 1998 *Mon. Not. R. Astron. Soc.* **299** 433 *Phys. Sin.* **59** 3618 (in Chinese) [鲍玉英、陈洛恩、张雄、彭朝阳、李怀珍 2010 物理学报 **59** 3618]
- [16] Sambruna R M 1997 *Astrophys. J.* **487** 536 [19] Liu Y, Zhang X, Zheng Y G, Wang X M, Bao Y Y 2007 *Acta Phys. Sin.* **56** 5558 (in Chinese) [刘云、张雄、郑永刚、王孝民、鲍玉英 2007 物理学报 **56** 5558]
- [17] Li W H, Yuan R 2009 *Acta Phys. Sin.* **58** 6671 (in Chinese) [李为虎、袁蓉 2009 物理学报 **58** 6671]
- [18] Bao Y Y, Chen L E, Zhang X, Peng Z Y, Li H Z 2010 *Acta Phys. Sin.* **59** 3618 (in Chinese) [鲍玉英、陈洛恩、张雄、彭朝阳、李怀珍 2010 物理学报 **59** 3618]
- [20] Yuan Y H, Fan J H, Huang Y 2008 *Chin. Phys. B* **17** 1526

Classification of Blazar objects based on Logistic regression model*

Tang Jie¹⁾ Wu Xue-Bing^{2)†}

1) (Department of Physics, Shaanxi University of Technology, Hanzhong 723001, China)

2) (Department of Astronomy, Peking University, Beijing 100871, China)

(Received 24 March 2011; revised manuscript received 16 May 2011)

Abstract

We collect 205 Blazars (142 BL Lacertae objects and 63 flat spectrum radio quasars (FSRQs)), and investigate the correlation between the redshift, the fluxes in the radio band (5 GHz), optical (V band), X-ray (1 keV), and the X-ray photon spectral index for the BL Lacertae objects and FSRQs by the correlation analysis and Logistic regression model. The results indicate that major influencing factors on Blazar object classification are the redshift, the fluxes in the radio band (5 GHz), and the X-ray photon spectral index. The derived equation is effective as a substantial criterion for Blazar objects classification, of which the accuracy is 91.2%. Our results strongly support the division of Blazar objects into BL Lacertae objects and FSRQs. We compare the properties of BL Lacertae objects and FSRQs from the fluxes in the optical (V band) and X-ray (1 keV), and find that BL Lacertae objects are not significantly different from FSRQs. The implications of these correlations, which support the evolutionary sequence of Blazar objects, are also discussed.

Keywords: Logistic, classification, Blazar objects

PACS: 98.54.Cm, 98.62.Ve, 02.50.-r, 98.52.Cf

* Project supported by the Key Program of the National Natural Science Foundation of China (Grant No. 11033001) and the National Basic Research Program of China (Grant No. 2007CB815405).

† Corresponding author. E-mail: wuxb@bac.pku.edu.cn