

利用小行星色指数区分S/C型小行星的方法*

叶嘉晖^{1,2†} 赵海斌^{1,3‡} 李 彬^{1,3}

(1 中国科学院紫金山天文台 南京 210008)

(2 中国科学院大学 北京 100049)

(3 中国科学院行星科学重点实验室 南京 210008)

摘要 利用多色测光得到的色指数在区分S/C型小行星过程中具有重要的参考意义, 由色指数空间主成分分析到的颜色参数 a 区分S/C类型是一种有效的方法. 采用 $a=0$ 分划的硬分割方法和对 a 作聚类分析法分别对SDSSMOC4和XSTPS-GAC的小行星多色数据样本进行分类, 结果表明: 两种方法对较完备的数据样本无明显差别, 且都与Carvano的S/C分类结果相符; $a=0$ 分划方法对样本空间变化敏感, 而 a 聚类分析方法更为稳定.

关键词 小行星: 普通, 方法: 数据分析

中图分类号: P185; 文献标识码: A

1 引言

小行星是围绕太阳运动的岩石或者金属天体, 它们是太阳系形成早期未能发生吸积的遗留物质, 因此仍然蕴含着太阳系原始星云的信息. S型和C型是最常见的小行星类型, 分别代表了小行星两种不同的光谱特征, 与它们的物质成分相关. S型常见于小行星主带内侧, 主要成分是硅酸盐岩石; C型常见于小行星主带中部和外侧, 主要成分是碳类物质. 物质成分的差异导致两类小行星的反照率不一样, S型小行星通常比同距离的相同大小的C型小行星亮. 从反射光谱的形状能够明显地区分两类小行星, 如图1所示. S型小行星光谱在小于 $0.7\mu\text{m}$ 的部分有很陡的上升斜率, 大于 $0.75\mu\text{m}$ 部分有中等至较强的吸收, 峰值在 $0.73\mu\text{m}$ 附近达到1.2左右. C型小行星光谱在小于 $0.55\mu\text{m}$ 部分有较弱至中等的紫外吸收, 大于 $0.55\mu\text{m}$ 部分平缓无特征. 其他类型的小行星光谱与S或C型小行星的光谱相似, 区别在于反照率大小和吸收带的强弱^[1].

2014-10-14收到原稿, 2014-11-14 收到修改稿

*国家自然科学基金项目 (11273067, 11178025)、小行星基金会、澳门科学技术发展基金项目 (095/2013/A3)及江苏省自然科学基金项目 (BK2011890) 资助

[†]yejiahui@pmo.ac.cn

[‡]meteorzh@pmo.ac.cn

小行星光谱观测耗费大量观测资源, 因此获得大样本连续光谱数据难度大, 而获得大样本多色测光数据则更可行^[2-3]. 利用多色测光数据得到小行星的色指数, 进一步对多个色指数开展主成分分析研究成为初步区分小行星类型的重要方法. 1984年Tholen使用主成分分析法研究了8色小行星巡天的结果, 从而总结出最早的被广泛使用的小行星分类方法^[4]. 2006年Roig和Gil-Hutton使用了主成分分析法在斯隆巡天的小行星中寻找候选V型小行星^[5]. 主成分分析法也能作为区分S/C型小行星的手段, 通过色指数构建出第1主成分, 定义为小行星的颜色参数 a , 它反映了光谱颜色偏红/蓝的特征, $a=0$ 是划分S和C型小行星的界线. Ivezić等^[6]和Roig等^[5]曾对斯隆巡天发布的不同批次数据作了同样的处理.

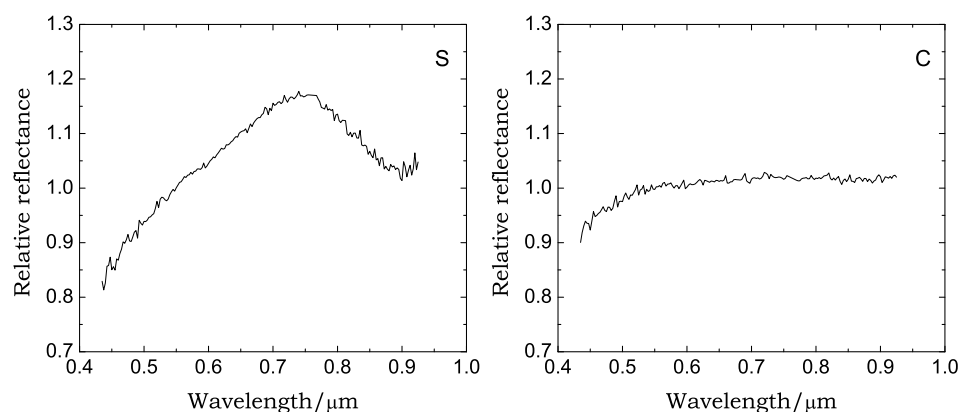


图1 典型S型和C型小行星在 $0.4\sim 1.0\ \mu\text{m}$ 内的反射光谱. 左图是S型, 右图是C型. 数值已归一到 $0.55\ \mu\text{m}$ 处^[1].

Fig. 1 The reflectance spectra of S type and C type asteroids within $0.4\sim 1.0\ \mu\text{m}$, normalized to unity at $0.55\ \mu\text{m}$. The left one is S type, and the right one is C type^[1].

以 $a=0$ 分划的区分方式对完备样本的分析结果与已有分类结果相符, 但是对有偏的样本数据库会得到不一致的结果, 对样本数量的敏感依赖性是该方法的固有缺点. 聚类分析法由于考虑了样本存在S/C类型的特征聚集现象, 其结果对样本数量的依赖性更小, 比 $a=0$ 分划的方式更合理, 其中K平均算法在已知分类的聚类分析法中是快速而有效的.

使用主成分分析和聚类分析, 目的是利用色指数区分S和C型小行星, 对继续区分小行星亚类有重要的参考意义. 本文应用这两种方法处理斯隆巡天中MOC4小行星数据和近地天体望远镜反银心方向小行星多色巡天的数据, 给出了有效的S/C分类, 并对结果进行了分析和讨论.

2 数据分析处理方法

2.1 主成分分析(Principal Component Analysis, PCA)

主成分分析是一种数学变换方法, 通过把一组相关的变量变换成另一组互不相关的变量, 新的变量按照方差大小排列, 第1变量具有最大的方差, 称为第1主成分, 方差次大的变量称为第2主成分, 以此类推. 变量的方差越大包含越多的信息, 因此第1主成分包含了原变量组的主要信息, 通过分析第1主成分来达到多维变量分析的降维目的. 在色指数研究中, 第1主成分反映了光谱的颜色特征.

通过多色巡天观测获取小行星色指数. 不同的观测项目使用数目不同的滤光片, 如8色小行星巡天使用了8个滤光片, 斯隆数字巡天使用了5个滤光片. 每个色指数在不同程度上反映了小行星的光谱信息, 相互之间有信息的重叠, 符合使用主成分分析的条件.

计算主成分的变换式为:

$$\mathbf{PC} = \mathbf{E}(\mathbf{F} - \mathbf{B}), \quad (1)$$

其中 $\mathbf{PC}$ 为所求的主成分, $\mathbf{F}$ 为色指数, $\mathbf{B}$ 为色指数的均值, $\mathbf{E}$ 为色指数协方差矩阵的特征向量.

2.2 K平均算法(K-means algorithm)

K平均算法属于硬聚类算法, 在已知类个数的前提下, 根据特征相似性划分数据集(类簇). 首先选择K个簇的初始平均值或中心, 对剩余的每个对象根据其与各个簇中心的距离, 将它赋给最近的簇, 然后重新计算每个簇的平均值. 这个过程不断迭代直至准则函数收敛.

使用K平均算法可以根据第1主成分 $\mathbf{PC}_1$ 把小行星分成S类和C类. S型和C型小行星是两类的主要成员, 聚集在两中心附近, 其他类型小行星也会被划分到特征相近的一类中. 由于类簇中心密集, 我们假定数据呈正态分布, 均值为 μ , 标准差为 σ , 区分依据是到两均值点的概率距离 d :

$$d = \frac{(x - \mu)^2}{\sigma^2}. \quad (2)$$

当 μ 和 σ 不再变化时, 所有小行星都能够唯一地划分到S或C类. 这里选择概率距离优于欧氏距离, 因为概率距离考虑了数据的聚集程度, 体现在标准差 σ 上.

2.3 星等归算

绝对星等计算采用Bowell等人在1985年提出的HG星等系统^[7], 可以表示为:

$$H(\alpha) = H - 2.5 \lg[(1 - G)\Phi_1(\alpha) + G\Phi_2(\alpha)], \quad (3)$$

其中 H 是小行星的绝对星等; α 是小行星的太阳相位角; $H(\alpha)$ 是一个与 α 有关的函数, 称为相位函数. G 值表征相位函数的斜率, 称为斜率参数. 不同的小行星有不同的 G 值, 反照率大的 G 较大, 反照率小的 G 较小. 对于未知 G 值的小行星, 通常取 $G=0.15$. 函数 Φ 表示为:

$$\Phi_i = W\Phi_{i1} + (1 - W)\Phi_{i2}, \quad i = 1, 2, \quad (4)$$

$$W = \exp[-90.56 \tan^2(\alpha/2)], \quad (5)$$

$$\Phi_{i1} = 1 - (C_i \sin \alpha) / (0.119 + 1.341 \sin \alpha - 0.754 \sin^2 \alpha), \quad (6)$$

$$\Phi_{i2} = \exp[-A_i \tan^{B_i}(\alpha/2)], \quad i = 1, 2, \quad (7)$$

其中 $A_1=3.332$, $B_1=0.631$, $C_1=0.986$, $A_2=1.862$, $B_2=1.218$, $C_2=0.238$.

3 结果与分析

3.1 斯隆数字巡天数据处理

斯隆巡天的移动天体数据星表(Sloan Digital Sky Survey Moving Object Catalog, SDSSMOC)中提供了小行星可见光波段高精度的色指数信息, 最新发布的是第4版数据, 观测历元至2007年3月. SDSSMOC4共计包含了104 539个小行星的220 101次观测¹.

图2为斯隆巡天 g 、 r 、 i 滤光片的响应曲线. 结合图1可见小行星的反射光谱特征区域落在 g 、 r 、 i 覆盖范围内. 为了与前人的工作结果保持一致, 我们使用 g 、 r 、 i 表示小行星在 g 、 r 、 i 波段内的绝对星等, 并省略单位mag. 根据Ivezić等^[6]的研究, 我们在选取小行星样本时选择色指数范围: $g-r$: $0\sim1.2$; $r-i$: $-0.2\sim0.6$, 共计选取103 641个小行星.

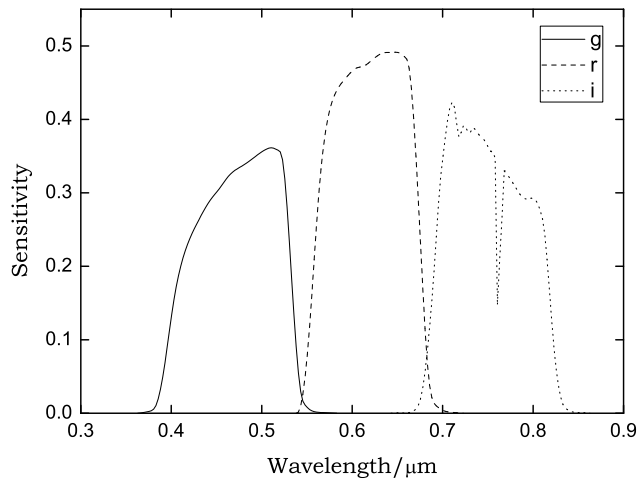


图2 斯隆巡天 g 、 r 、 i 滤光片的响应曲线

Fig. 2 The filter profiles of SDSS g , r , and i wavebands

图3为SDSSMOC4的 $(g-r)$ - $(r-i)$ 色指数分布, 从图中可以看出明显的双峰结构. 双峰结构主要由两大类型小行星的分布造成, 如图4所示. S型和C型的聚集区域分别对应图3中的右上峰和左下峰. 小行星光谱在 $0.3\sim1\mu\text{m}$ 内有两个主要吸收特征: 一个小于 $0.55\mu\text{m}$, 另一个大于 $0.7\mu\text{m}$. r 波段处于两个吸收带之间, 所以小行星的 $g-r$ 和 $r-i$ 色指数能够反映光谱的变化情况, S型和C型小行星色指数的差异是造成双峰分布的原因.

根据(1)式得到的小行星颜色参数 a 为:

$$a = 0.9696(g - r) + 0.2448(r - i) - 0.6263. \quad (8)$$

实际上 a 是色指数在特征向量方向上的投影, 如图3中虚线所示, 根据主成分分析的原理, 数据的方差在该方向上最大. a 反映了小行星的光谱形态或颜色信息, a 越大表示光谱颜色偏红, a 越小表示光谱颜色偏蓝. Ivezić等^[6]在2001年分析了最早发布的斯隆数据, 根据6 150个小行星得到的 a 值为:

$$a = 0.89(g - r) + 0.45(r - i) - 0.57. \quad (9)$$

¹ <http://sbn.psi.edu/pds/resource/sdssmoc.html>

2006年Roig和Gil-Hutton根据SDSSMOC3中17 957 个小行星得到的 a 值为^[5]:

$$a = 0.9285(g - r) + 0.3713(r - i) - 0.6204. \quad (10)$$

可见不同版本斯隆数据的 a 值随着小行星样本数量的增加而变化, $g-r$ 系数会变大而 $r-i$ 系数则变小.

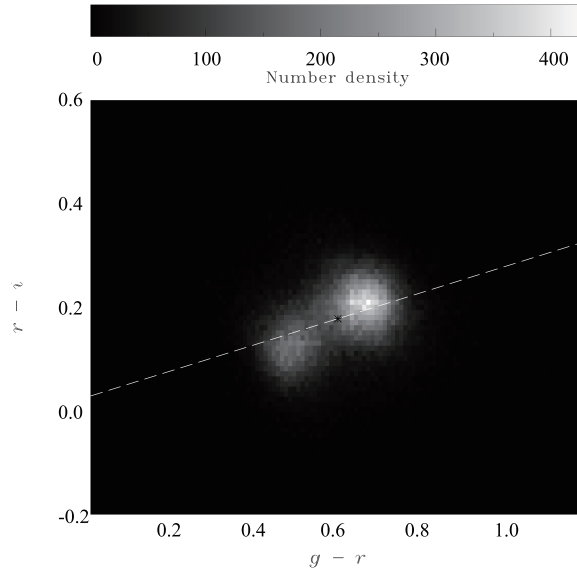


图3 SDSSMOC4小行星的 $(g-r)-(r-i)$ 色指数分布

Fig. 3 The distributions of SDSSMOC4 asteroids' $(g-r)-(r-i)$ color indexes

根据Ivezić等^[6]的分析结果, 我们也把 $a=0$ 作为区分S型和C型小行星的依据. 对S型小行星 $a>0$; 对C型小行星 $a<0$. 需要说明的是小行星并不只有S和C两种类型, a 值的大小只说明小行星的颜色类S或C型, 小行星的类型还需要结合反照率、物质成分等才能最终确定. 聚类分析选择 $a=0$ 划分结果中两类的均值和标准差作为K平均算法的初始均值和标准差. 迭代后得到: $\mu_S=0.0747$ 、 $\sigma_S=0.0659$ 和 $\mu_C=-0.1192$ 、 $\sigma_C=0.0592$. 两类一正一负的均值点与 $a=0$ 分划的结果一致, 标准差很小说明数据很聚集. 图5为拟合得到的小行星样本关于参数 a 的高斯分布, 其中纵轴 $f(a)=\frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{(a-\mu)^2}{2\sigma^2}]$ 表示颜色参数为 a 的小行星数量的相对概率密度.

为了比较两种分类方式的效果, 我们用Carvano等^[8]的分类结果作了对比². Carvano分类是最新的使用广泛的小行星分类系统. 表1是两种分类结果与Carvano系统S/C型小行星数目的比较. 由比较结果可见两种分类方式对S和C型小行星的区分非常符合已知的分类结果. 对于数据较完备的样本, 如SDSSMOC4, 以 $a=0$ 和聚类划分的结果没有明显差别.

² <http://sbn.psi.edu/pds/resource/sdsstax.html>

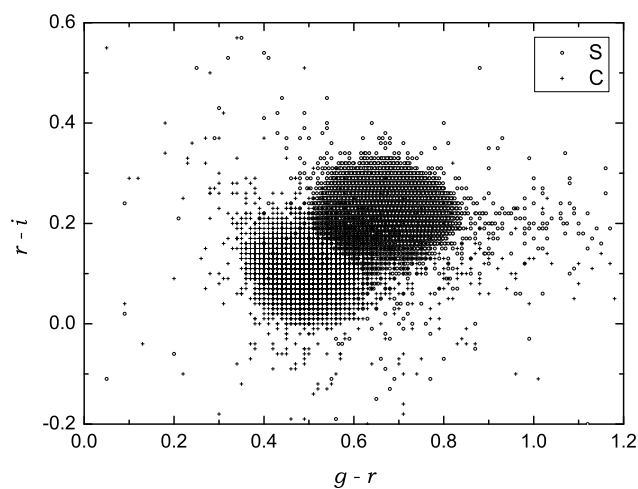


图 4 S型和C型小行星的 $(g-r)$ - $(r-i)$ 色指数分布

Fig. 4 The distributions of S type and C type asteroids' $(g-r)$ - $(r-i)$ color indexes

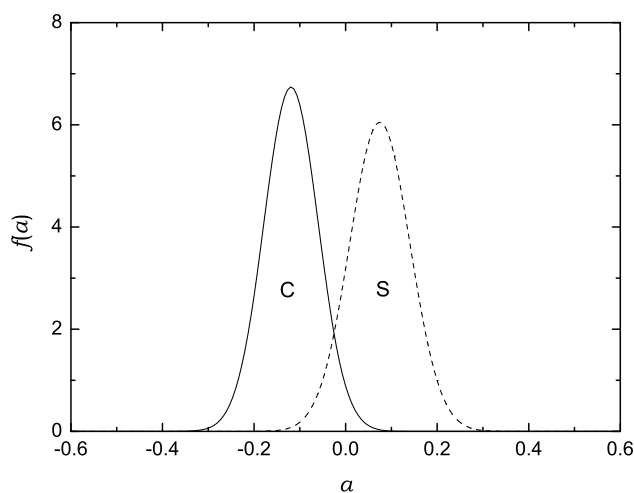


图 5 SDSSMOC4小行星 a 参数高斯分布. 左右两个分布分别对应C型和S型小行星.

Fig. 5 The Gaussian distributions of “ a ” derived from the SDSSMOC4 data. The left and right distributions correspond to C type and S type asteroids, respectively.

表 1 SDSSMOC4两种分类结果与Carvano S/C型小行星的数目比较
Table 1 The comparison of our results of SDSSMOC4 derived by PCA and cluster analysis methods respectively with those of the Carvano's method

	$a \geq 0$	$a < 0$	K-means_S	K-means_C
Carvano_S	15190 (96.3%)	576 (3.7%)	15623 (99.1%)	143 (0.9%)
Carvano_C	664 (3.8%)	16964 (96.2%)	1544 (8.8%)	16084 (91.2%)

主成分分析的 a 值依赖于样本, 由图4可见两类小行星的分布有交集, 以数据均值 $a=0$ 作为划分界线属于硬分割, 这种划分方式没有考虑交集部分. 聚类分析考虑了数据有聚集和交集现象, 以两个正态分布模拟双峰分布, 所以更符合实际情况. 当样本包含的S型和C型小行星数目有很大差别时, $a=0$ 不能准确区分两种类型, 但是聚类分析依然有效. 我们随机选择了Carvano分类中不同数量的S和C型小行星组成新样本作分类处理. 结果表明以 $a=0$ 划分的结果对不同的样本空间有波动, 而聚类分析的划分结果相对稳定, 聚类分析更适合有选择性样本的数据处理. 因此为了都能够使用两种分类方法, 我们应该选择数据完备的样本. 表2和表3是两种不同情况的对比结果.

表 2 7 000 个S型、3 000 个C型小行星使用两种划分方法与Carvano分类的数目对比
Table 2 The comparison of the results of 7000 S type and 3000 C type asteroids derived by PCA and cluster analysis methods respectively with those of the

	Carvano's method			
	$a \geq 0$	$a < 0$	K-means_S	K-means_C
Carvano_S	6526 (93.2%)	474 (6.8%)	6934 (99.1%)	66 (0.9%)
Carvano_C	55 (1.8%)	2945 (98.2%)	177 (5.9%)	2823 (94.1%)

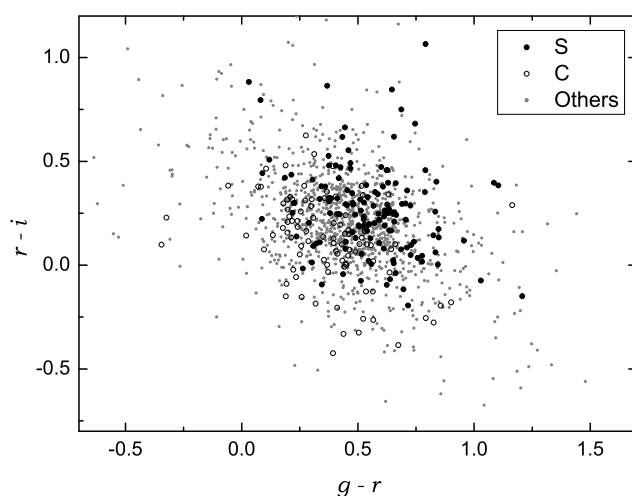
表 3 8 000 个S型、2 000 个C型小行星使用两种划分方法与Carvano分类的数目对比
Table 3 The comparison of the results of 8000 S type and 2000 C type asteroids derived by PCA and cluster analysis methods respectively with those of the

	Carvano's method			
	$a \geq 0$	$a < 0$	K-means_S	K-means_C
Carvano_S	6699 (83.7%)	1301 (16.3%)	7862 (98.3%)	138 (1.7%)
Carvano_C	26 (1.3%)	1974 (98.7%)	87 (4.3%)	1913 (95.7%)

3.2 盱眙近地天体望远镜反银心方向测光巡天数据处理

盱眙近地天体望远镜反银心方向测光巡天(Xuyi Schmidt Telescope Photometric Survey of the Galactic Anti-center, XSTPS-GAC)由北京大学科维理天文与天体物理研究所和紫金山天文台合作完成, 观测历元为2009年10月至2010年3月、2010年9月至2011年3月, 历时2 yr共13个月. 在以银河系反银心方向为中心、赤经 $3^{\text{h}} \sim 9^{\text{h}}$ 、赤纬 $-10^{\circ} \sim +60^{\circ}$ 约6 000平方度的天区以及覆盖仙女座大星云、三角座风车星云约600平方度的扩展天区开展了CCD成像测光观测. 观测使用了斯隆巡天g、r、i滤光片, 每个观测夜只在一个波段观测, 其中g波段观测42 d (2009年10月至2011年3月), r波段观测37 d (2010年9月至2011年3月), i波段观测47 d (2009年10月至2011年3月). 有关XSTPS-GAC的数据处理流程, 在Zhang等的研究中有详细介绍^[9-10].

因为各小行星的g、r、i波段的观测数据是在2 yr内的不同观测历元获取的, 小行星的相位角变化、小行星的自转效应均会引起小行星光度的变化, 这些原因最终造成了色指数的弥散. 图6为XSTPS-GAC的1 173个小行星的 $(g-r)$ - $(r-i)$ 色指数分布, 其中实心圆和空心圆分别表示S型和C型小行星. 从图中无法识别色指数的双峰结构, S型和C型小行星的聚集区域没有分离.

图 6 XSTPS-GAC小行星的 $(g-r)-(r-i)$ 色指数分布Fig. 6 The distributions of XSTPS-GAC asteroids' $(g-r)-(r-i)$ color indexes

尽管色指数没有明显的聚集分布,但仍然能够使用主成分分析和聚类分析.主成分分析得到的 a 值为:

$$a = 0.9168(g-r) - 0.3993(r-i) - 0.3598. \quad (11)$$

聚类分析的结果为: $\mu_S=0.1468$ 、 $\sigma_S=0.1497$ 和 $\mu_C=-0.1837$ 、 $\sigma_C=0.1117$. XSTPS-GAC色指数的方差比SDSSMOC4大.图7为拟合的S和C型小行星 a 参数的分布曲线.

根据图6,对于XSTPS-GAC数据未能明确地区分出S/C两类小行星,主要是由于观测数据的方差大所致.表4为两种分类结果与Carvano系统S/C型小行星数目的比较.

表 4 XSTPS-GAC两种分类结果与Carvano S/C型小行星的数目比较

Table 4 The comparison of our results of XSTPS-GAC derived by PCA and cluster analysis methods respectively with those of the Carvano's method

	$a \geq 0$	$a < 0$	K-means_S	K-means_C
Carvano_S	81 (65.3%)	43 (34.7%)	91 (73.4%)	33 (26.6%)
Carvano_C	27 (31.4%)	59 (68.6%)	42 (48.8%)	44 (51.2%)

4 结论和讨论

研究色指数分类方法的目的是对已有多色测光观测的但无光谱观测的小行星作初步类型判定,以及为确定小行星亚类提供参考信息.本文通过对小行星色指数使用主成分分析法构建颜色参数 a ,并利用 $a=0$ 分划和 a 聚类分析来区分S/C型小行星,得出结论如下:

(1) 分析得到了SDSSMOC4的 $g-r$ 、 $r-i$ 色指数的双峰分布, S/C型分类结果表

明 $a=0$ 分划和 a 聚类分析方法对数据较完备的样本的区分无明显差别, 与Carvano的分类结果相符合.

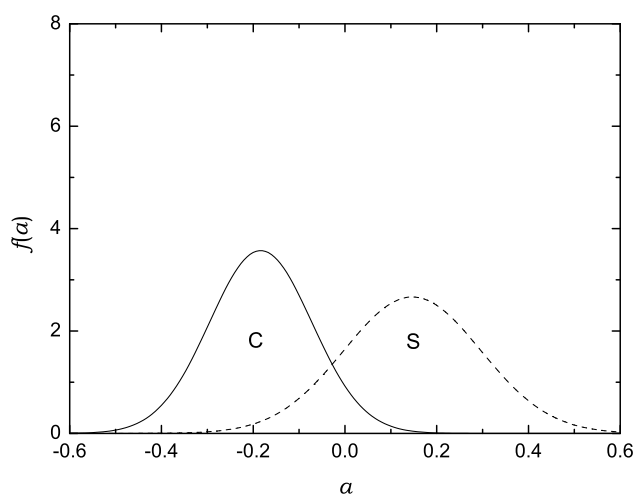


图 7 XSTPS-GAC小行星 a 参数高斯分布. 左右两个分布分别对应C型和S型小行星.

Fig. 7 The Gaussian distributions of “ a ” derived from the XSTPS-GAC data. The left and right distributions correspond to C type and S type asteroids, respectively.

(2) 由于 a 值的大小依赖于样本空间, 而聚类分析考虑了数据的聚集分布和交集情况, 所以 $a=0$ 分划的结果对样本数量敏感, 而聚类分析的结果则更稳定.

(3) 对XSTPS-GAC的小行星测光数据分析表明, 对于小行星的色指数方差大的情况, 仍然可以使用两种区分方法, 只是数据弥散会弱化区分结果, 甚至导致无法区分.

$a=0$ 分划和 a 聚类分析法的有效性得益于 g 、 r 、 i 波段覆盖了小行星光谱最重要的两个吸收特征, 所以即使只有两个色指数, 仍然能够与S/C类型有很强的相关性, 从而达到区分的目的. 该方法在具有较完备的SDSSMOC4样本中的应用得到了证实, 而XSTPS-GAC的数据由于观测方式决定了色指数弥散大, 因此未来需要获取多波段同时观测的数据.

致谢 衷心感谢紫金山天文台王歆研究员为本文的研究方法和结果提出的宝贵意见.

参考文献

- [1] Bus S J, Binzel R P. *Icar*, 2002, 158: 146
- [2] 李雪婧. *天文学报*, 2010, 51: 384
- [3] Li X J. *ChA&A*, 2010, 35: 117
- [4] Tholen D J. *Asteroid Taxonomy from Cluster Analysis of Photometry*. Tucson: University of Arizona, 1984
- [5] Roig F, Gil-Hutton R. *Icar*, 2006, 183: 411
- [6] Ivezić Ž, Tabachnik S, Rafikov R, et al. *AJ*, 2001, 122: 2749
- [7] Lagerkvist C I, Williams I P. *A&AS*, 1987, 68: 295
- [8] Carvano J M, Hasselmann P H, Lazzaro D, et al. *A&A*, 2010, 510: A43
- [9] Zhang H H, Liu X W, Yuan H B, et al. *RAA*, 2013, 13: 490

[10] Zhang H H, Liu X W, Yuan H B, et al. RAA, 2014, 14: 456

Methods of Asteroid S/C Type Classification Using Color Index

YE Jia-hui^{1,2} ZHAO Hai-bin^{1,3} LI Bin^{1,3}

(1 Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008)

(2 University of Chinese Academy of Sciences, Beijing 100049)

(3 Key Laboratory of Planetary Sciences, Chinese Academy of Sciences, Nanjing 210008)

ABSTRACT Color indexes are important to classify the S/C type asteroids. The asteroids' color parameter a reduced from principal component analysis of color indexes is a key parameter of classification. Two methods, dividing by the parameter a directly and the cluster analysis by parameter a , are applied in asteroids' S/C classification of Sloan Digital Sky Survey (SDSS) and Xuyi Schmidt Telescope Photometric Survey of the Galactic Anti-center. The SDSSMOC4 shows a bimodal distribution in the $g-r$ vs $r-i$ diagram. The results show that these two methods are effective and consistent in the case of large samples, well in accordance with the Carvano's consequences. But the result divided by parameter a directly is unstable for different subsets, while the cluster analysis is more stable. Both methods still can be used even the variances of the color indexes are large, but the results of classification are not obvious.

Key words asteroids: general, methods: data analysis