

CmII 奇宇称光谱能级的 PCA研究*

Pattern Recognition Applied to the Energy Levels of the CmII Odd-parity Spectrum

李殷青

郭进**

Li Yinqing

Guo Jin

(广西师范大学分析测试中心 桂林市育才路 3号 541004)

(Analysing & Testing Centre, Guangxi Normal Univ., 3 Yucailu, Guilin, Guangxi, 541004, China)

摘要 利用模式识别的主成分分析法 (PCA-Principal Component Analysis) 对 CmII 奇宇称光谱的电子组态进行分类, 得出 $5f^8 7p$ $5f^7 6d 7s$ $5f^7 6d^2$ 各组态间的分类判据。结果表明: 选择适当的模式识别方法和空间投影, 寻找组态间的分类判据, 可以有效地对原子光谱的电子组态进行分类和预报, 是原子光谱分析的一种有用的方法。

关键词 CmII 奇宇称光谱 电子组态 主成分分析

中图分类号 O 562.3

Abstract The principal component analysis, one of pattern recognition methods, is applied to classification of electronic configuration of the odd-parity spectrum of curium (CmII) and some criteria for classification of configurations among $5f^8 7p$ $5f^7 6d 7s$ $5f^7 6d^2$ are obtained. The results show that electronic configurations of atomic spectrum can be assigned and predicted effectively by classification criteria obtained by both pattern recognition and space mapping, which are useful for atomic spectrum analysis.

Key words CmII odd-parity spectrum, electronic configuration, principal component analysis

原子光谱的归属通常是通过研究能级差别的规律来确定的。激发态的电子组态可通过谱线的相对强度、同位素位移和塞曼效应等数据来指认, 也可通过量子理论的计算 (如 Hartree-Fock 方法) 来指认, 实验观测与理论计算的结合在很大的范围上获得了相当完整的原子光谱分类。但由于光谱的复杂性, 特别是较高能级的激发态往往出现谱线的混合。因此, 在实验观测和理论计算上很难对这些高能级激发态的原子光谱进行有效的指认。然而, 随着计算机应用技术的发展而出现一种人工智能技术——化学模式识别技术, 为原子光谱的指认提供了一种有效的方法。

Peterson 曾利用 KNN (K Nearest Neighbors, K 个最邻近法) 和实验数据对 CmI、H、U、U II 等原子光谱进行研究, 并对未确定的高激发态光谱线进行了指认^[1~4], 也利用神经网络算法中的对神经网络方法 (CPN) 对 CmII 光谱进行过研究^[5], 这些研

究都取得了较好的效果。虽然如此, 但对光谱指认的准确率只有 80%~90%, 究其原因, 主要是模式识别方法的选取。最近, 刘洪霖等人^[6~7]利用反向传播方法对 CmI、U 光谱重新指认, 准确率达 100%。鉴于模式识别方法分类的有效性, 本文利用模式识别中的主成分分析法对 CmII 奇宇称光谱线进行分类研究, 以探索光谱电子组态的组成规律。

1 原理方法

主成分分析法是一种多元统计分析技术, 其目的是将数据降维, 以排除众多信息中相互重叠的信息。它是将原变量进行转换, 将原变量进行线性组合成新变量, 这些变量尽可能地表征原变量的数据结构而不丢失信息。

主成分分析法的原理是对多维空间作 Karhunen-Loeve 变换, 即将样本通过一正交线性变换成两两正交的新变量

$$T = XP,$$

T 为主成分 (也称得分矢 score) 变量, X 为样本特征变量矩阵, P 为正交变换矩阵, 由相互正交的矢量 p_i 构成, 主成分 t_i 可由下式得到

$$t_i = Xp_{.i}$$

1998-07-06收稿。

* 国家自然科学基金 (59861001)、教育部优秀年轻教师基金 (1999-5)、广西科学配套基金 (桂科配 9912005) 和广西教育厅科学基金 (1998-169) 资助项目。

** 广西大学物理系, 南宁市西乡塘路 10号, 530004 (Dept. of Phys., Guangxi Univ., 10 Xixiangtanglu, Nanning, Guangxi, 530004, China)

具体计算步骤如下: 设有 m 个变量的 N 个样本观测值, 数据矩阵 (标准化) 为

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

(1) 计算协方差阵 Σ 的估计值

$$\hat{\Sigma} = (S_{ij}) = S,$$

其中

$$S_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j),$$

$$\bar{x}_{ij} = \frac{1}{n} \sum_{k=1}^n x_{ki}.$$

(2) 用 Jacobi 方法求 S 的特征根 λ_i 和相应的特征

矢量 $l_i (i = 1, \dots, m)$.

(3) 写出各主成分表达式

$$t_k = \sum_{j=1}^m l_{jk} x_{kj},$$

以 t 为空间矢量构成一多维空间, 并在某二维面上投影以寻找可视分类图。

本文的 CmII 奇宇称光谱线的能级数据取自文献 [8], 以光谱线的能级 $E (10^{-4} \text{ cm}^{-1})$, 同位素位移 $S (\text{cm}^{-1})$, 总角动量量子数 J 和 Landé 因子 g 作为特征参量, 文献 [8] 中的 79 个已知的 CmII 奇宇称光谱线分属电子组态: $5f^8 7p$, $5f^7 6d 7s$, $5f^7 6d^2$, 并在以下的分类中分别以 1 类, 2 类, 3 类表示。计算工作采用自编程序在 586 微机上进行。

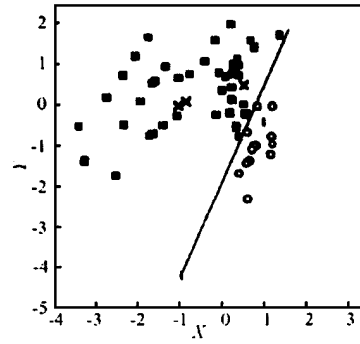
2 结果与讨论

一般的光谱理论认为: 对于原子光谱, 其谱线的电子组态可以由谱线的同位素位移 S 来确定, 且同一组态的谱线有相同的同位素位移。实际上, 同一组态谱线的同位素位移是不尽相同的, 例如, 已知 CmII 奇宇称光谱的 $5f^8 7p$, $5f^7 6d 7s$, $5f^7 6d^2$ 组态同位素位移的范围分别为: $-1.022 \sim -0.659$, $-0.757 \sim -0.429$, $-1.017 \sim -0.8$, 可以看出不同组态的同位素位移有部分是重叠的。因此, 仅靠同位素位移来判定谱线的归属是相当困难的, 特别是随着谱线的能级的增加, 组态混合增加, 对谱线的指认就更为困难。

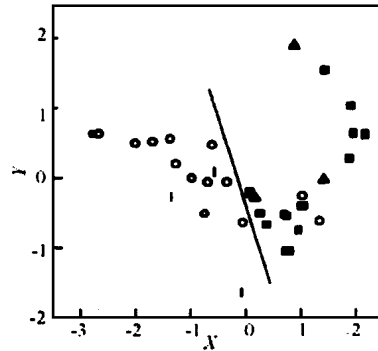
Peterson^[5] 也认为对原子光谱指认的困难主要在于同位素位移重叠区域的高能级谱线。

利用模式识别方法可以有效地帮助对谱线的指认。图 1 为对 79 条已知的 CmII 奇宇称光谱线的能级电子组态进行的模式识别分类图。从图中可以看

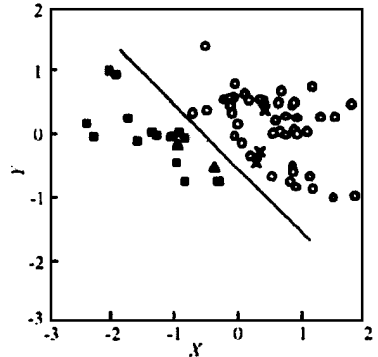
出: 除了 $5f^8 7p$ 组态有两条谱线落到 $5f^7 6d^2$ 组态区域外 (图 1b), $5f^8 7p$, $5f^7 6d 7s$, $5f^7 6d^2$ 各组态已被明显的分离, 分类的准确率为 97.5%。



$$\begin{aligned} a) X &= -2.21 + 0.69E - 2.66S + 0.12J - 0.71g \\ Y &= 2.43 - 0.15E + 4.32S + 0.41J - 0.71g \end{aligned}$$



$$\begin{aligned} b) X &= -3.42 - 1.03E - 6.74S + 0.06J + 0.13g \\ Y &= 9.43 - 1.04E + 6.72S + 0.02J - 0.24g \end{aligned}$$



$$\begin{aligned} c) X &= 1.94 - 0.17E + 4.38S + 0.32J + 0.0001g \\ Y &= -2.33 + 0.79E + 1.26S - 0.09J + 1.15g \end{aligned}$$

图 1 CmII 奇宇称光谱线的电子组态分类: $5f^8 7p$ (1 类), $5f^7 6d 7s$ (2 类), $5f^7 6d^2$ (3 类). \cdot, \times, \triangle (预报点), 图 1a 中 \circ 为 1 类, \blacksquare 为 2 类; 图 1b 中 \circ 为 1 类, \blacksquare 为 3 类; 图 1c 中 \circ 为 2 类, \blacksquare 为 3 类。

Fig. 1 The classifications of electronic configuration of the CmII odd-parity spectrum: $5f^8 7p$ (Class 1), $5f^7 6d 7s$ (Class 2), $5f^7 6d^2$ (Class 3). \cdot, \times, \triangle (prediction). \circ Class 1, \blacksquare Class 2 (in Fig. 1a); \circ Class 1, \blacksquare Class 3 (in Fig. 1b); \circ Class 2, \blacksquare Class 3 (in Fig. 1c).

表 1 若干 CmII 奇宇称光谱的指认

Table 1 The assignment for some electronic configurations of CmII odd-parity spectrum

No.	E/cm^{-1}	S/cm^{-1}	J	g	图 1 中标记 Mark in Fig. 1	本文指认 Assignment	实验指认 ^[8] Reference ^[8]
1	15 559. 133	- 0. 659	9/2	1. 650	×	$5f^7 6d7s$	$5f^7 6d7s$
2	17 013. 379	- 0. 800	9/2	1. 633	△	$5f^7 6d^2$	$5f^7 6d^2$
3	18 205. 773	- 0. 634	9/2	1. 585	×	$5f^7 6d7s$	$5f^7 6d7s$
4	29 925. 300	- 0. 974	11/2	1. 425	△	$5f^7 6d^2$	$5f^7 6d^2$
5	30 745. 000	- 0. 990	1/2	- 1. 050	△	$5f^7 6d^2$	$5f^7 6d^2$
6	31 933. 755	- 0. 635	11/2	1. 274	×	$5f^7 6d7s$	$5f^7 6d7s$
7	31 947. 675	- 0. 859	7/2	1. 478		$5f^8 7p$	$5f^8 7p$
8	33 265. 480	- 0. 830	11/2	1. 382		$5f^8 7p$	$5f^8 7p$
9	36 083. 640	- 0. 958	1/2	1. 245		$5f^8 7p$	$5f^8 7p$

从图 1 各组态间的分界线可确定相应分类判据

(1) $5f^8 7p$ 与 $5f^7 6d7s$ 组态间的分类判据为 (图 1a):

$$- 26. 84 + 4. 66 \times 10^{-4} E - 28. 43S - 0. 41J - 5. 43g > 0 \quad \text{为 } 5f^8 7p \text{ 组态 (1类);}$$

(2) $5f^8 7p$ 与 $5f^7 6d^2$ 组态间的分类判据为 (图 1b):

$$1. 77 - 4. 01 \times 10^{-4} E - 11. 25S + 0. 19J + 0. 09g < 0 \quad \text{为 } 5f^8 7p \text{ 组态 (1类);}$$

(3) $5f^7 6d7s$ 与 $5f^7 6d^2$ 组态间的分类判据为 (图 1c):

$$0. 50 + 1. 86 \times 10^{-4} E + 16. 71S + 0. 67J + 3. 43g > 0 \quad \text{为 } 5f^7 6d7s \text{ 组态 (2类);}$$

本文根据这些判据对若干“未知”谱线进行指认。

例如, 对文献 [8] 中的 9 条没有参与模式识别分类的能级进行了谱线归属指认, 结果列于表 1

从图 1 及表 1 可以看出, 对 9 条“未知”谱线的指认结果与实验相符。从分类判据看出谱线的能级 E 同位素位移 S 总角动量子数 J 和 Landé 因子 g 对谱线的电子组态分类的贡献。虽然 E, S, J, g 共同影响对谱线的电子组态的分类, 但 E 和 S 是分类的主要因素。在一些特殊情况下, 可以很方便地对谱线进行指认。例如, 对 $5f^8 7p$ 和 $5f^7 6d^2$ 组态的谱线, 组态间的分类判据 $1. 77 - 4. 01 \times 10^{-4} E - 11. 25S + 0. 19J + 0. 09g < 0$ 为 $5f^8 7p$ 组态反之则为 $5f^7 6d^2$ 组态。对于低能级谱线, 分类判据 > 0 , 为 $5f^7 6d^2$ 组态; 对高能级谱线, 分类判据 < 0 , 为 $5f^8 7p$ 组态。据此, 对已知 $5f^8 7p$ 和 $5f^7 6d^2$ 组态的谱线的能级的排列结果为: $5f^8 7p$ 组态的能级范围 27 065. 085~ 39 336. 155, $5f^7 6d^2$ 组态的能级范围 14 830. 15~ 32 923. 88 从所

列的能级范围证实了分类判据的分析。对处于两组态的能级重叠区的谱线, 则须通过 E, S, J, g 等 4 因素的综合影响分析, 即通过分类判据来判定。

总之, 选择适当的模式识别方法和空间投影, 寻找组态间的分类判据, 可以有效地对原子光谱的电子组态进行分类和预报, 不失为原子光谱分析的一种有效方法。

参考文献

- Peterson K L, Parsons M L. Spectral classification using pattern-recognition techniques I. Feasibility with hydrogen as a model system. Phys Rev, 1978, A17: 261.
- Peterson K L, Anderson D L. Spectral classification using pattern-recognition techniques II. Application to curium energy levels. Parsons M L Phys Rev, 1978, A17: 270.
- Lewis R V, Peterson K L. Spectral classification of uranium I energy levels using pattern-recognition techniques. Phys Rev, 1987, A35: 1119.
- Lewis R V, Peterson K L. Spectral classification of UIII energy levels using pattern-recognition techniques. Phys Rev, 1988, A38: 3773.
- Peterson K L. Classification of CmII energy levels using counter propagation neural network. Phys Rev, 1990, A41: 2457.
- 曹晓卫, 刘洪霖, 陈念贻. Cm I 奇宇称光谱线的能级的模式识别研究. 物理化学学报, 1996, 5 (12): 400.
- 曹晓卫, 刘洪霖, 陈念贻. UI 奇宇称光谱线的能级的化学模式识别方法研究. 科学通报, 1997, 42 (21): 2284.
- Worden E F, Conway J G, Blaise J. Energy levels of the second spectrum of Curium CmII. J Opt Soc Am, 1986, 3 (8): 1092.

(责任编辑: 黎贞崇)