

模式识别技术在不同来源甘草药材分类中的应用*

卓林, 朱尔一**, 王巧娥, 王小如

(厦门大学化学化工学院化学系 现代分析科学教育部重点实验室, 厦门 361005)

摘要 目的:建立不同来源甘草药材的模式识别分类方法。**方法:**通过 HPLC 快速获得 19 个不同来源甘草药材样品的色谱指纹图谱, 色谱条件固定相为 Kromasil KR100-5 C₁₈ (150 mm × 4.6 mm), 流动相为乙腈-0.05% 三氟乙酸溶液, 梯度洗脱为乙腈 20% (0 min) → 40% (5 min) → 50% (10 min), 流速为 0.8 mL · min⁻¹, 检测波长为 254 nm, 从每个色谱指纹图中提取出 325 个特征变量, 采用聚类分析对甘草药材样品进行相似性评价, 并应用 PLS 方法建立起不同植物种、不同种植方式甘草的 2 个分类回归模型。**结果:**采用分类模型对不同植物种、不同种植方式甘草进行判别, 2 个分类回归模型分类准确率均为 100%。**结论:**此方法可用于不同来源甘草药材的分类研究, 本文以 HPLC 指纹图谱结合化学计量学知识, 为不同中药样品的质量评价提供了一种分类方法。

关键词: HPLC; 指纹图谱; 聚类分析; PLS; 甘草药材

中图分类号: R917

文献标识码: A

文章编号: 0254-1793(2006)08-1080-05

Application of pattern recognition technology in classification of Radix et Rhizoma Glycyrrhizae of different sources*

ZHUO Lin, ZHU Er-yi**, WANG Qiao-e, WANG Xiao-ru

(The Key Laboratory for Analytical Science of the MOE, Department of Chemistry,

College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China)

Abstract Objective: To establish a method for pattern recognition classification on Radix et Rhizoma Glycyrrhizae of different sources. **Methods:** HPLC system was applied to obtain the chromatograms of the 19 Radix et Rhizoma Glycyrrhizae samples. The Kromasil KR100-5 C₁₈ (150 mm × 4.6 mm) column was used. A mixture of acetonitrile 0.05% trifluoroacetic acid solution was used as the mobile phase with a flow rate of 0.8 mL · min⁻¹ and the gradient mode was acetonitrile 20% (0 min) → 40% (5 min) → 50% (10 min). The detection wavelength was 254 nm. The 325 characteristic variables were obtained from each chromatogram. The clustering analysis was applied to evaluate the similarity of the chromatograms. Then partial least-squares regression method was applied to build the prediction models of classification which were used to distinguish Radix et Rhizoma Glycyrrhizae of different species and cultivation modes. **Results:** The classification accuracy of the two prediction models of classification was 100% for Radix et Rhizoma Glycyrrhizae of different sources. **Conclusion:** This method is practicable in the pattern recognition classification on Radix et Rhizoma Glycyrrhizae of different sources. The developed technique could be further applied to the classification of different traditional Chinese medicine (TCM).

Key words: HPLC; fingerprint; clustering analysis; PLS; Radix et Rhizoma Glycyrrhizae

中药的质量评价一直是羁绊中药研究与应用的难点和重点问题, 我国有些传统中药及制剂由于缺乏科学、先进、可行的质量标准, 难以有效地控制在质量。

近 10 年来, 模式识别方法已逐渐深入地应用于

中药品质评价, 划分中药材质量等级, 真伪、优劣鉴别等。其中, 化学模式识别方法已经成为该领域卓有成效的方法^[1-3], 但有关偏最小二乘法 (PLS 方法) 应用于不同中药分类的报道甚少, 本文通过 HPLC 快速获得 19 个不同来源甘草药材样品的色

* 福建省自然科学基金 (C021006); 福建 GAP 关键技术研究 (2002Y024)

** 通讯作者 Tel: (0592) 2184510; E-mail: ryzhu@xmu.edu.cn

谱指纹图,并从每个色谱指纹图中采集 325 个特征变量,进行聚类分析研究和 PLS 模型分类研究,得到较为满意的分类结果。

1 材料和仪器

19 个甘草药材样品(表 1),由内蒙古亿利科技实业股份有限公司“内蒙梁外甘草 GAP 研究与开发”项目部采集并由中国医学科学院药用植物研究所林寿全教授鉴定。

表 1 甘草药材样品来源

Tab 1 Source of Radix et Rhizoma Glycyrrhizae samples

| 编号 (No.) * | 产地 (sources) | 种植方式 (cultivation modes) | 采收日期 (collect - untime) |
|---------------|-------------------------------------|--------------------------------|-------------------------------|
| 1 | 新疆布尔津(Bu'erjing, Xinjiang) | 人工(cultivated) | 2001-08 |
| 2 | 新疆和静(Hejing, Xinjiang) | 人工(cultivated) | 2001-08 |
| 3 | 新疆布尔津(Bu'erjing, Xinjiang) | 野生(wild) | 2001-08 |
| 4 | 新疆和静(Hejing, Xinjiang) | 野生(wild) | 2001-08 |
| 5 | 河北承德(Chengde, Hebei) | 人工(cultivated) | 2000-09 |
| 6 | 河北石家庄(Shijiazhuang, Hebei) | 人工(cultivated) | 2000-09 |
| 7 | 山西侯马(Houma, Shanxi) | 人工(cultivated) | 2000-08 |
| 8 | 黑龙江安达(An'da, Heilongjiang) | 野生(wild) | 2002-08 |
| 9 | 吉林白城(Baicheng, Jilin) | 人工(cultivated) | 2000-08 |
| 10 | 吉林大安(Da'an, Jilin) | 野生(wild) | 2002-08 |
| 11 | 内蒙古元宝山(Yuanbaoshan, Nei Monggol) | 人工(cultivated) | 1998-08 |
| 12 | 敖汗旗赤峰(Chifeng, Ao'hanqi) | 野生(wild) | 1998-08 |
| 13 | 陕西延安(Yan'an, Shaanxi) | 人工(cultivated) | 2001-08 |
| 14 | 宁夏盐城(Yancheng, Ningxia) | 野生(wild) | 1998-08 |
| 15 | 甘肃庆阳(Qingyang, Gansu) | 人工(cultivated) | 1999-08 |
| 16 | 新疆哈密(Hami, Xinjiang) | 野生(wild) | 2001-08 |
| 17 | 新疆巴楚(Bachu, Xinjiang) | 野生(wild) | 2001-08 |
| 18 | 新疆和田(Hetian, Xinjiang) | 野生(wild) | 2001-08 |
| 19 | 新疆阿克苏(A'kesu, Xinjiang) | 野生(wild) | 2001-08 |

*1~15. 甘草(*Glycyrrhiza uralensis* Fisch.); 16~19. 胀果甘草(*Glycyrrhiza inflata* Bat.)

甘草酸对照品,内蒙古亿利科技实业股份有限公司李秉经理提供,纯度为 80%,由 50% 乙醇溶液配制,浓度为 250 $\mu\text{g} \cdot \text{mL}^{-1}$ 的对照品溶液。

Agilent 1100 HPLC 仪(美国安捷伦公司,包括输液泵、脱气装置、自动进样装置和 DAD 检测器); MIKRO 22R 冷冻离心机(德国 Hettich 公司); SK3200LH 超声波清洗器(上海科导超声仪器有限公司); LD-101B 型方形调节式中药切片机(中国温岭市大海药材机械厂); LD-06A 型高速中药粉碎机(中国温岭市大海药材机械厂)。

2 实验方法

2.1 色谱分析条件^[4] 色谱柱:Kromasil KR100-5 C₁₈(150 mm × 4.6 mm);柱温:室温;进样量:15 μL ;

流动相:乙腈-0.05% 三氟乙酸溶液;梯度洗脱:乙腈:20%(0 min)→40%(5 min)→50%(10 min);流速:0.8 $\text{mL} \cdot \text{min}^{-1}$;检测:DAD 检测器,254 nm;参考波长:500 nm;波长宽度:16 nm。

2.2 样品溶液制备 样品经切片机切片、粉碎机粉碎,过 60 目筛后置干燥器中备用。所有样品在使用前均在烘箱中 60 $^{\circ}\text{C}$ 下烘 15 h,取出置干燥器中冷却至室温备用。取甘草粉末样品 0.25 g,精密称定,置于 25 mL 量瓶中,加入 80 $^{\circ}\text{C}$ 热水 20 mL,超声(功率 133 W,频率 59 kHz)提取 1 h,冷却至室温,定容;12070 $\text{r} \cdot \text{min}^{-1}$ 离心 20 min,提取上层离液,过 0.45 μm 微孔滤膜后即得。

2.3 精密度试验 取 1 号甘草样品溶液连续进样 5 次,甘草酸峰面积的 RSD 为 1.2%,其他峰面积的 RSD 均小于 1.5%,甘草酸保留时间的 RSD 为 1.3%,其他峰保留时间的 RSD 小于 2.0%。

2.4 重复性试验 制备 1 号甘草样品溶液 5 份,进样测定,甘草酸峰面积的 RSD 为 1.3%,其他峰面积的 RSD 均小于 2.0%,甘草酸保留时间的 RSD 为 1.5%,其他峰保留时间的 RSD 小于 2.0%。

2.5 稳定性试验 取 1 号甘草样品溶液,每隔 1 d 进样 1 次,连续 5 d,甘草酸峰面积的 RSD 为 1.5%,其他峰面积的 RSD 均小于 2.0%,甘草酸保留时间的 RSD 为 1.5%,其他峰保留时间的 RSD 小于 2.0%。

2.6 样品测定 19 个甘草药材样品按照“2.2”方法制备样品溶液并进样,即得相应的 HPLC 色谱图,典型的色谱图或数据如 1 号甘草药材的色谱图,见图 1-B 所示,本实验所得色谱图均未扣除仪器噪音和本底。

3 数据处理与结果讨论

3.1 特征变量的提取

本实验中甘草药材色谱图由 HPLC 数据采集系统获得,将其导入到 EXCEL 软件中,可得 2400 多个二维数据点,为方便数据处理,考虑对色谱图进行压缩校正,提取出相应的特征变量。由于甘草酸是甘草药材的主要成分和特征指标,通常可用来评价甘草药材的质量^[5],而每张色谱图中的甘草酸(最大峰)的保留时间值略有差异,在特征变量提取前需要对整个保留时间尺度做一个校正,使甘草酸于同一时间出峰。

方法:(1). 对每个甘草药材色谱图的每个二维数据点,保持纵轴信号强度值不变,将横轴保留时间值均乘上校正系数 $m(m = \bar{i}/t, \bar{i}$ 是所有甘草药材色谱图中甘草酸保留时间的平均值,本实验取接近值

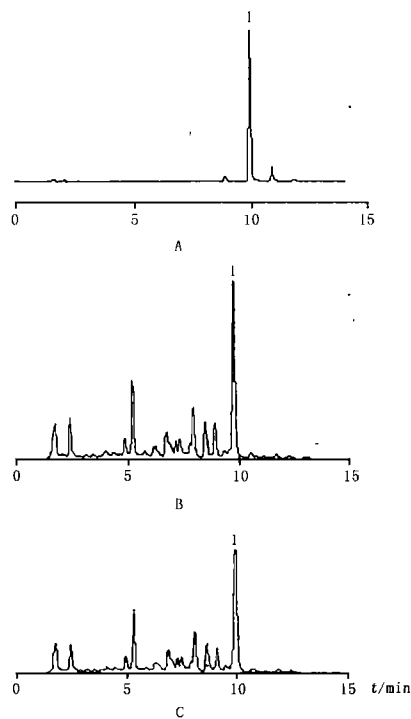


图1 250 $\mu\text{g} \cdot \text{mL}^{-1}$ 对照品溶液(A)、1号新疆布尔津甘草色谱图经压缩校正前(B)及压缩校正后(C)HPLC色谱图

Fig 1 HPLC chromatograms of 250 $\mu\text{g} \cdot \text{mL}^{-1}$ reference substance solution(A), the original HPLC chromatogram(B) and the corrected HPLC chromatogram(C) of the No. 1 Bu'erjing Xinjiang *Glycyrrhiza uralensis* Fisch.

1. 甘草酸(glycyrrhizic acid)

为9.9 min, t 为所选甘草药材色谱图中甘草酸的保留时间值,使甘草药材色谱图中甘草酸(最大峰)的保留时间为统一值(9.9 min);(2).以校正后的横轴保留时间值作为参照,从0 min开始,至14 min结束,计算每0.04 min内相应数据点纵轴信号强度的平均值,计算出325个平均值,每个平均值即为一个特征变量;3.从0 min开始,至14 min结束,每隔0.04 min对应一个特征变量作图,即得压缩校正后的甘草药材色谱图,校正后每个色谱图含有325个特征变量,1号甘草药材样品压缩校正前后色谱图如图1所示,可以看出无明显差异。

19个甘草药材样品的色谱图经压缩校正后,可得到维数为 325×19 的特征变量数据矩阵 X ,供聚类分析和PLS建模使用。

3.2 聚类分析研究

结合特征变量数据矩阵 X ,采用相关系数法^[6]对甘草药材样品进行相似性评价,聚类分析结果如图2所示,19个不同来源甘草药材样品聚为3个类别,9~18号甘草药材样品聚为第一类别,表明9~18号甘草药材样品较为相似;1~8号甘草药材样品

聚为第二类别,表明1~8号甘草药材样品较为相似;19号甘草药材样品独聚为第三类别,3个类别体现了甘草药材不同产地、不同植物种以及不同种植方式的品质差异。就种植方式而言,第二类别中1及2号新疆人工甘草药材、3及4号新疆野生甘草药材、5及6号河北人工甘草药材各自先聚成类,以及第一类别中11号内蒙人工甘草药材、13号陕西人工甘草药材和15号甘肃人工甘草药材为同一小类,表明种植方式的不同对甘草药材品质的相似性评价有重要影响;虽然新疆的4种甘草药材与4种胀果甘草药材存在明显的差异,但是新疆产地的4种胀果甘草药材在聚类图中分布较为零散,野生甘草药材与人工甘草药材在聚类图中也未完全分开,这可能与中药化学成分的复杂性与模糊性有关。

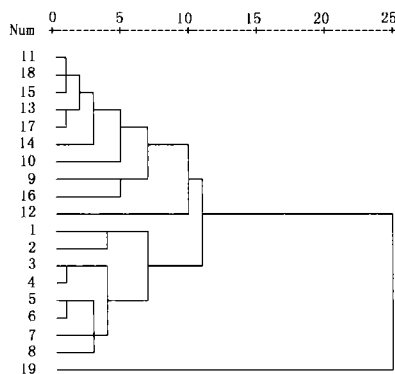


图2 聚类分析树状图

Fig 2 Dendric drawing of clustering analysis

3.3 分类模型研究

鉴于植物来源、生长方式对甘草药材品质的相似性评价均有影响,而用直接观察指纹图谱方法,发现不同植物种、不同种植方式的甘草药材的指纹图谱十分相似,难以分辨,故采用PLS方法对不同植物种、不同种植方式甘草药材做进一步的模型分类研究。PLS法是一种数据“软”建模的稳健统计方法,通常用于数据探索性分析,尤其当解释变量个数多、样本量少时很有效^[7],本实验有19个甘草药材样品,325个数据变量,可采用PLS方法分析并建立回归模型: $Y = X\beta$, β 为回归系数,根据回归模型对甘草药材样品的预报值来判别甘草药材类别。

方法:先根据甘草药材样品的不同类别构造目标变量矩阵 Y ,对于 Y 中元素,如果样本属于第一类别,则取1,若属第二类则取2,将 Y 与变量矩阵 X 进行PLS正交分解,由第一、第二正交分量构成判别平面分类图^[8],并采用使预报残差平方和值(Prediction sum of squares, PRESS值)最低或接近最低的

隐变量个数建立回归模型, PRESS 值越小, 表示模型的预报能力越强, 根据回归模型对甘草药材样品的预报值来判别甘草药材类别, 如果预报值接近 1, 则样品属于第一类别, 如果预报值接近 2, 则样品属于第二类别^[9,10]。

3.3.1 不同植物种甘草药材的模型分类

取甘草为第一类别, 胀果甘草为第二类别, 构造不同植物种甘草药材的目标变量矩阵 Y_1 (表 2), 将 Y_1 与数据矩阵 X 进行 PLS 回归分析, 由第一、第二正交分量 PLS - t_1 、PLS - t_2 , 得出不同植物种甘草药材的判别分类图 (图 3), 采用使 PRESS 值为 0.368 (最低值) 的 8 个隐变量建立回归模型, 并计算出各个样品的模型预报值 Y_1' (表 2)。

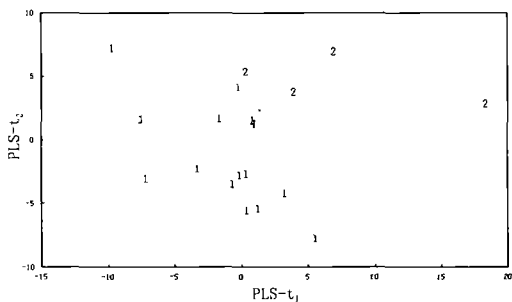


图 3 不同植物种甘草药材的 PLS 模式识别分类图

Fig 3 The discrimination classification figure for Radix et Rhizoma Glycyrrhizae of different species

1. 甘草 (*Glycyrrhiza uralensis* Fisch.) 2. 胀果甘草 (*Glycyrrhiza inflata* Bat.)

表 2 不同植物种甘草药材的模型分类

Tab 2 The model classification on Radix et Rhizoma Glycyrrhizae of different species

| 样品编号 (sample No.) | 目标变量 Y_1 (dependent - variable Y_1) * | 模型预报 Y_1' (prediction - result Y_1') | 判别类别 (discriminant result) |
|-------------------------|--|---|----------------------------------|
| 1 | 1 | 1.021 | 1 |
| 2 | 1 | 1.046 | 1 |
| 3 | 1 | 0.959 | 1 |
| 4 | 1 | 1.041 | 1 |
| 5 | 1 | 0.996 | 1 |
| 6 | 1 | 0.948 | 1 |
| 7 | 1 | 1.023 | 1 |
| 8 | 1 | 0.963 | 1 |
| 9 | 1 | 0.975 | 1 |
| 10 | 1 | 0.986 | 1 |
| 11 | 1 | 1.025 | 1 |
| 12 | 1 | 0.99 | 1 |
| 13 | 1 | 1.037 | 1 |
| 14 | 1 | 1.018 | 1 |
| 15 | 1 | 0.988 | 1 |
| 16 | 2 | 2.03 | 2 |
| 17 | 2 | 1.984 | 2 |
| 18 | 2 | 1.987 | 2 |
| 19 | 2 | 1.984 | 2 |

*1. 甘草 (*Glycyrrhiza uralensis* Fisch.); 2. 胀果甘草 (*Glycyrrhiza inflata* Bat.)

观察图 3, 甘草与胀果甘草分类较为清晰; 观察 Y_1' , 1 ~ 15 号甘草药材样品预报值均接近 1, 判断为第一类别甘草, 16 ~ 19 号甘草药材样品预报值接近 2, 判断为第二类别胀果甘草, 判别类别与实际类别一致, 可见不同植物种甘草药材的回归模型分类准确率为 100%。

3.3.2 不同种植方式甘草药材的模型分类

取野生甘草药材为第一类别, 人工甘草药材为第二类别, 同法构造不同种植方式甘草药材的目标变量矩阵 Y_2 (表 3), 将 Y_2 与数据矩阵 X 进行 PLS 回归分析, 由第一、第二正交分量 PLS - t_1 、PLS - t_2 同样得出不同种植方式甘草药材的判别分类图 (图 4), 采用使 PRESS 值为 0.105 (最低值) 的 8 个隐变量建立回归模型, 并计算出各个样品的模型预报值 Y_2' (表 3)。

表 3 不同种植方式甘草药材的模型分类

Tab 3 The model classification on Radix et Rhizoma Glycyrrhizae of different cultivation modes

| 样品编号 (sample NO.) | 目标变量 Y_2 (dependent - variable Y_2) * | 模型预报 Y_2' (prediction - result Y_2') | 判别类别 (discriminant result) |
|-------------------------|--|---|----------------------------------|
| 1 | 2 | 2.002 | 2 |
| 2 | 2 | 1.970 | 2 |
| 3 | 1 | 1.034 | 1 |
| 4 | 1 | 0.968 | 1 |
| 5 | 2 | 2.033 | 2 |
| 6 | 2 | 2.030 | 2 |
| 7 | 2 | 1.987 | 2 |
| 8 | 1 | 0.976 | 1 |
| 9 | 2 | 2.010 | 2 |
| 10 | 1 | 1.002 | 1 |
| 11 | 2 | 2.003 | 2 |
| 12 | 1 | 1.001 | 1 |
| 13 | 2 | 1.935 | 2 |
| 14 | 1 | 1.008 | 1 |
| 15 | 2 | 2.019 | 2 |
| 16 | 1 | 0.994 | 1 |
| 17 | 1 | 0.994 | 1 |
| 18 | 1 | 1.037 | 1 |
| 19 | 1 | 0.998 | 1 |

*1. 野生甘草药材 (wild Radix et Rhizoma Glycyrrhizae); 2. 人工种植甘草药材 (cultivated Radix et Rhizoma Glycyrrhizae)

观察图 4, 人工甘草药材与野生甘草药材分类较为清晰; 观察 Y_2' , 其中 3, 4, 8, 10, 12, 14 以及 16 ~ 19 号甘草药材样品预报值接近 1, 判断为第一类别野生甘草药材, 其余样品预报值接近 2, 判断为第二类别人工种植甘草药材, 判别类别同样与实际类别一致, 可见不同种植方式甘草药材的回归模型分类准确率同样为 100%。

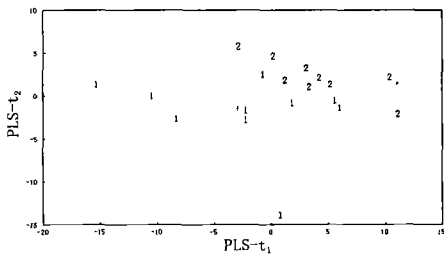


图4 不同种植方式甘草药材的 PLS 模式识别分类图

Fig 4 The discrimination classification figure for Radix et Rhizoma Glycyrrhizae of different cultivation modes

1. 野生甘草药材(wild Radix et Rhizoma Glycyrrhizae)
2. 人工种植甘草药材(cultivated Radix et Rhizoma Glycyrrhizae)

4 讨论

本文通过 HPLC 分析方法快速获得 19 个不同来源甘草药材样品的色谱指纹图谱,对色谱图进行压缩校正,提取出 325 个特征变量用于数据处理和综合评价,采用聚类分析和 PLS 建立分类模型方法对问题进行分析研究,聚类分析结果表明,新疆产地的 4 种甘草药材与 4 种胀果甘草药材属于不同的类别,存在较明显的差异,种植方式的不同对甘草药材品质的相似性评价也有重要影响,但是新疆产地的 4 种胀果甘草药材在聚类图中分布较为零散,野生甘草药材与人工甘草药材在聚类图中也未完全分开。而采用 PLS 法建立分类模型的研究表明,不同植物种、不同种植方式的甘草药材较为清晰地地区分开来(见图 3、图 4),2 个回归模型的分类准确率均为 100%,表明 PLS 法能适合样本个数较少,变量个数较多的数据解析,由于不同类别甘草药材色谱指纹图十分相似,难以用肉眼观察法进行分辨,而用 PLS 法建立分类模型的方法能很准确地将其分辨,因此,PLS 法建立分类模型的方法能为不同中药样品的质量评价提供较为可靠的分类方法。

致谢:本实验得到了厦门开元区高新技术研究开发基地王文慎经理、邱招钗老师的帮助,在此表示感谢!

参考文献

- 1 MA Ying-li(马英丽), ZHAO Huai-qing(赵怀清), TIAN Zhen-kun(田振坤). Studies on chemical pattern recognition of quality assessment of *Radix Astragali*(黄芪质量的化学模式识别研究). *Chin Tradit Herb Drugs*(中草药), 2003, 34(5):460
- 2 CHEN Min-jun(陈闽军), WU Yong-jiang(吴永江), FAN Xiao-hui(范晓辉). Identification of the habitat of *Ligusticum chuanguixiong* with chromatographic fingerprinting(色谱指纹图谱技术用于鉴别中药川芎产地). *China J Chin Mater Med*(中国中药杂志), 2003, 28(7):606
- 3 ZHANG Xing-hui(张兴辉), SHI Li-fu(石力夫). Studies on different origin of the fruits of *Ligustrum lucidum* by chemical pattern recognition(不同产地中药女贞子的化学模式识别研究). *Pharm J Chin PLA*(解放军药理学学报), 2004, 20(6):447
- 4 WANG Qiao-e(王巧娥), WANG Wen-shen(王文慎), ZHANG Ji-shu(张吉树), et al. Study on optimal harvest time of semi-wild *Glycyrrhiza uralensis* at Liangwai, inner Mongolia(梁外半野生甘草最佳采收期研究). *J Chin Med Mater*(中药材), 2004, 27(4):235
- 5 ChP(中国药典). 2000. Vol I(一部):65
- 6 WANG Xi(王玺), WANG Wen-yu(王文宇), ZHANG Ke-rong(张克荣). Discussion on the studies of the similarity of the HPLC fingerprints of traditional Chinese medicine(中药 HPLC 指纹图谱相似性研究探讨). *J Shenyang Pharm Univ*(沈阳药科大学学报), 2003, 20(5):362
- 7 JIANG Hong-wei(蒋红卫), XIA Jie-lai(夏结来). Partial least square and its application(偏最小二乘回归及其应用). *J Fourth Mil Med Univ*(第四军医大学学报), 2003, 24(3):280
- 8 ZHANG Yuan(章元), ZHU Er-yi(朱尔一), ZHUANG Zhi-xia(庄峙厦). Classification study by pattern recognition on the relationship between the trace elements in human hair and sex(人发微量元素与性别关系的模式识别分类研究). *Chem J Chin Univ*(高等学校化学学报), 1998, 19(7):1055
- 9 ZHU Er-yi(朱尔一), YANG Peng-yuan(杨芃原). Technology and Application of Chemometrics(化学计量学技术及应用). Beijing(北京): Science Press(科学出版社), 2001. 92
- 10 ZHU Er-yi(朱尔一). A kind of PLS method suitable to deal with the fingerprinting data of Chinese medicine(一种适合于处理中药指纹图谱数据的偏最小二乘法). *Comput Appl Chem*(计算机与应用化学), 2005, 22(8):639

(本文于 2005 年 9 月 19 日收到)