

# 中药药性物质基础研究中数学建模思路探讨

王晓燕<sup>1,2\*</sup>

(1. 山东中医药大学药学院, 济南 250355; 2. 山东中医药大学理工学院, 济南 250355)

**[摘要]** **目的:**探讨数学建模策略在中药药性物质基础研究中的应用,并尝试建立一套客观可重复的建模思路与方法。**方法:**从数据库建立、数据预处理、建模策略选择和评价标准制定 4 个关键步骤,讨论中药药性特征标记识别模型的建立。**结果:**从 64 种建模策略中选出 17 种判别良好的模型,继而考察其可视化表达能力选出 4 种适宜模型,最后经统计模拟,选出 PLS 作为目前最适宜模型。**结论:**中药药性与其发挥作用的物质基础之间难以找到现成的公式或规律来表达二者关联,可用穷举策略,经数据库建立、数据预处理、建模策略选择和评价 4 步筛选适宜的中药药性特征标记识别模型。

**[关键词]** 数学建模; 药性; 物质基础; 穷举建模策略

**[中图分类号]** R287 **[文献标识码]** A **[文章编号]** 1005-9903(2013)15-0337-03

**[doi]** 10.11653/syfy2013150337

## Study on Mathematical Modeling in Material Foundation of Traditional Chinese Medicine

WANG Xiao-yan<sup>1,2\*</sup>

(1. College of Pharmaceutical, Shandong University of Traditional Chinese Medicine (TCM), Ji'nan 250355, China;  
2. College of Science and Technology, Shandong University of TCM, Ji'nan 250355, China)

**[Abstract]** **Objective:** To discuss the application of mathematical modeling strategies in material foundation of traditional Chinese medicine (TCM), and try to establish a set of objective modeling methods. **Method:** From the database establishment, data preprocessing, modeling strategy choosing and model's evaluations the most appropriate recognition model for CHMP-markers was choose. **Result:** Seventeen kind of modeling strategies from 64 was selected, and 4 kind was chosen in view of the visual expression ability, and PLS was selected as the most appropriate one via statistical simulation. **Conclusion:** It is difficult to find ready-made formula or law express the relation between the relation of Chinese medicine medicinal properties and material base, exhaustive method can be used to choose the most suitable model, database establishment, data preprocessing, modeling strategy choosing and model's evaluations are the 4 key steps.

**[收稿日期]** 20121126(007)

**[基金项目]** 国家重点基础研究发展计划项目(2007CB512601)

**[通讯作者]** \*王晓燕, 博士生, 讲师, 从事中药质量控制与资源研究, E-mail: sdnwxy@126.com

[5] Jeffrey W Seifert. Data mining: An overview [R]. CRS Report RL31798, 2004.  
[6] 薛为民, 陆玉昌. 文本挖掘技术研究[J]. 北京联合大学学报: 自然科学版, 2005, 19(4): 59.  
[7] 吕婷, 姜友好. 文本挖掘在生物医学领域中的应用及系统工具[J]. 中华医学图书情报杂志, 2010, 19(4): 56.  
[8] 李立, 周奇, 郑光, 等. 基于文本挖掘技术分析中成药、西药对慢性胃炎的治疗规律[J]. 中国实验方剂

学杂志, 2011, 17(24): 228.  
[9] 郭洪涛, 郑光, 赵静, 等. 基于文本挖掘分析甲型 H1N1 流感的中医药治疗特色[J]. 世界科学技术——中医药现代化, 2011(5): 772.  
[10] 谭勇, 杨静, 赵宁, 等. 利用文本挖掘技术探索中西药治疗慢性乙型肝炎的用药规律[J]. 中国实验方剂学杂志, 2011, 17(24): 232.

[责任编辑 邹晓翠]

[ **Key words** ] mathematical modeling; Chinese medicine medicinal properties; material base; exhaustive method

数学作为中药由文字描述为主过渡到定量研究的重要手段<sup>[1-2]</sup>,已经在中药现代研究中得到广泛运用,如金锐等<sup>[3]</sup>用模糊数学探讨药性判别模式,汤金波等<sup>[4]</sup>用统计学探讨中药性-效关联,此外数学在质量评定、性味归经等方面也有应用<sup>[5-7]</sup>。但以上研究均未建立一套各家公认的研究思路。本文以寻找中药药性与其物质基础之间关系的中药性特征标记识别模型的建立为例,探讨建模策略应用,并尝试建立一套客观、可重复、可推广的建模思路与方法。从中药整体药性出发,利用指纹图谱技术(以 HPLC 为例)分离中药物质成分,并拟据图谱数据建立统计模型,以找到可表征中药药性的各图谱标记作为中药药性特征标记(中医基础理论认为,寒与凉、温与热仅程度不同无本质之分,为简化推理过程本文所指药性仅指寒热两性)。

**1 材料与方**

**1.1 数据库建立** 据《本草》、《本草纲目》等权威记载,并参照《中国药典》2010 年版,挑选 60 味药性明确的中药(寒、热各 30 味),从各药材市场统一购买(表 1),经山东中医药大学鉴定教研室李峰教授鉴定均为道地药材。对全部样本采用同一条件<sup>[9]</sup>进行 HPLC 分析。

每味样品平行提取 2 份/次,各测 2 次,终得数据 4 次。在 Spec Align 2.4<sup>[10]</sup> 中用差值算法依次导入各次数据,以消除噪声与误差,保证结果具相同保留时间值,再取均值。得每样品 190 ~ 400 nm 共 6 524 个保留时间点信号强度值,见图 1。

**1.2 数据预处理** 为进一步建模需要,数据需进行压缩:以波长为特征变量,对应信号值为吸光度,药性寒、热为因变量。利用 PCA 分析,发现同味药相邻波长下数据具强相关性,故可挑选一些特定波长代表邻近波长大部分信息,终选 12 个代表性波长(图 1)。同味药不同波长下出峰位置基本一致,但峰值略有不同,故将 12 个波长下数据缩至 [0, 1],再取均值。

为抑制噪声及系统误差,在 SpecAlign 中经数据分箱、谱线平滑、基线校正、峰强度标准化、谱峰联配、谱峰提取与量化,最终导出 96 个保留时间点信号强度值(即潜在寒热药性特征标记)作为建模的原始数据见图 2。

**1.3 建模策略及评价标准** 欲识别寒热药性特征

表 1 60 种中药样品

名称	药性	采样地区	名称	药性	采样地区
天冬	寒	贵州	肉桂	大热	广西
瓜蒌	寒	山东	附子	大热	四川
知母	寒	河北	吴茱萸	热	贵州
金银花	寒	山东	干姜	热	四川
墨旱莲	寒	山东	高良姜	热	广东
生地黄	寒	河南	仙茅	热	四川
紫草	寒	新疆	荜拔	热	海南
淡竹叶	寒	浙江	白胡椒	热	海南
防己	寒	浙江	杜仲	温	四川
侧柏叶	寒	山东	天南星	温	山东
甘遂	寒	陕西	陈皮	温	广东
秦皮	寒	辽宁	厚朴	温	四川
关龙胆	寒	辽宁	木瓜	温	安徽
芦荟	寒	云南	羌活	温	四川
海藻	寒	山东	藁本	温	四川
蒲公英	寒	山东	威灵仙	温	江苏
大黄	寒	甘肃	麻黄	温	内蒙古
黄柏	寒	四川	白芥子	温	安徽
栀子	寒	江西	川芎	温	四川
黄连	寒	重庆	淫羊藿	温	陕西
大青叶	寒	河北	木香	温	云南
瞿麦	寒	山东	法半夏	温	四川
豨莶草	寒	江苏	苍术	温	江苏
地肤子	寒	山东	檀香	温	广东
车前子	微寒	江西	草豆蔻	温	海南
川贝母	微寒	四川	甘松	温	四川
柴胡	微寒	河南	补骨脂	温	四川
络石藤	微寒	江苏	细辛	温	辽宁
葛根	凉	山东	红花	温	河南
薄荷	凉	江苏	炙元胡	温	浙江

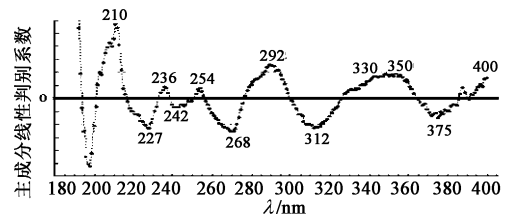


图 1 基于紫外光谱的主成分-线性判别

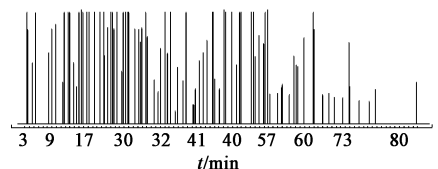


图 2 数据预处理后 60 味中药 HPLC

标记,首先要正确判别药性寒热。为寻找中药药性

特征标记识别适宜统计模型,本研究选用贝叶斯分类器、学习算法、函数、规则分类器 4 类共 64 种可能有效的方法进行建模,评价标准:①用判别正确率、真(假)阳性率及综合得分等指标,通过组内回代、外推预测、交叉验证选择判别效果好的加入候选集;②其中有判别函数表达式,各预测变量(潜在药性特征标记)对应系数可量化表达的加入适宜集。

## 2 结果与分析

据评价标准①综合得分  $\geq 0.8$  的 17 种加入候选集。考察标准②仅 SMO, PLS, Ridge Logistic, Bayesian Logistic Regression 4 种可加入适宜集。

4 种模型识别正确率难分轩轻,继续考察样品

潜在寒热药性特征标记实际差值(图 3)及各模型判别系数标识的理论图谱(图 4)。可见 PLS 理论图谱与实际差值图最相似,故认为 PLS 是目前基于 HPLC 的中药药性特征标记最适宜识别模型。

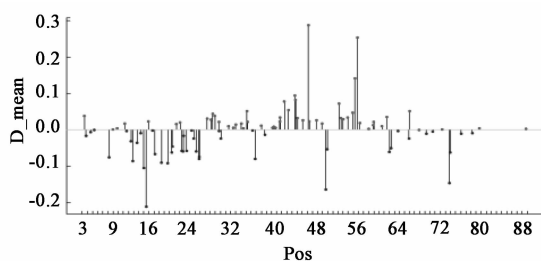


图 3 60 味植物药实际差值图谱

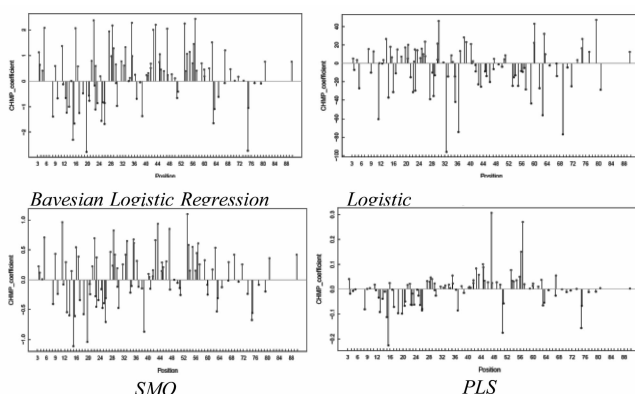


图 4 由 4 种较适宜模型判别系数绘制的中药药性特征标记理论图谱比较

## 3 讨论

在创新研究中很多问题难以找到求解规律或公式,故穷举建模策略是必然之选。本研究旨在探求中药药性与其物质基础之间的关系,根据样品 HPLC 数据,运用穷举策略,构建了数据库建立、数据预处理、统计模式识别建模策略和模型比较评价 4 个关键步骤,筛选出 PLS 作为中药药性特征标记统计识别模型。本思路客观可重复,可作为其他中药研究领域数学建模思路的借鉴。

本研究因经费所限,仅选用 60 味中药样本,扩大样本含量以评价所构建的统计模式识别模型的良好性和可推广性,并在应用中加以检验和修正,使之逐步完善,是进一步工作的重点。

### [参考文献]

[1] 史道华,陈鹭颖,郑炜. 中药药效研究的数学思考[J]. 中国实验方剂学杂志,2007,13(8):66.  
 [2] 赵汉青,江依法,葛鑫,等. 基于中药药性加权计算的中医方剂评价方法[J]. 中国实验方剂学杂志,2011,17:303.  
 [3] 金锐,张冰,刘小青. 中药寒热药性表达模糊评价模

式理论与实验研究[J]. 中西医结合学报,2012,10:1106.

[4] 汤金波,帅真,许仕杰. 基于数学思维构建中药“性-效关联”模型[J]. 世界科学技术——中医药现代化,2011,6:1099.  
 [5] 孙奕,王景明. 模糊数学在淫羊藿免疫药理学研究中的应用[J]. 数理医药学杂志,1998,11(4):350.  
 [6] 于莲波. 用模糊数学原理评估中药材质量等级[J]. 烟台师范学院学报,1997,13(3):9.  
 [7] 王咏梅,马红,刘苏中. 方剂配伍的模糊数学特性研究[J]. 中国实验方剂学杂志,2000,6(6):59.  
 [8] 欧阳兵,王振国,李峰,等. 中药四性“性-效-物质三元论”假说及其论证[J]. 山东中医药大学学报,2008,32(3):182.  
 [9] 王晓燕,李峰. 20 种植物药多糖水解成分 HPLC 指纹图谱与寒热药性关系的 SVM 分析研究[J]. 山东中医药大学学报,2012,36(5):439.  
 [10] Wong J W H, Cagney G, Cartwright H M. SpecAlign-processing and alignment of mass spectra datasets[J]. Bioinformatics, 2005, 21(9):2088.

[责任编辑 邹晓翠]