

基于支持向量机的黄连饮片产地识别研究

陶梦琳¹, 顾文涛¹, 汪子青¹, 侯珂惠¹, 崔书盛¹, 唐道超¹, 秦娜¹, 张大永^{2,3}, 万军^{1*}

1. 西南交通大学生命科学与工程学院、电气工程学院, 四川 成都 610031

2. 四川新荷花中药饮片股份有限公司, 四川 成都 610036

3. 国家中医药管理局中药炮制技术重点实验室, 四川 成都 610036

摘要: **目的** 建立基于支持向量机的黄连饮片产地区分识别模型。**方法** 采集4个产地多批黄连样本, 量化外部特征值(包括形状、气味、味道等), 并按《中国药典》2010年版方法测定内部特征值(包括水分、总灰分、醇浸出物, 指标成分表小檗碱、黄连碱、巴马汀、小檗碱质量分数), 在Matlab 7.0平台进行数据降维和融合, 建立黄连产地区分模型。**结果** 单独分析各项数据不能较好区分各产地黄连饮片, 而采用支持向量机建模后所有特征叠加识别率达到97.1%, 能准确区分各黄连饮片产地, 内外特征的高识别率说明各特征子集间具有一定的互补性, 可综合辨识不同产地黄连饮片的差异性。**结论** 建立的基于支持向量机的识别模型, 实现产地的区分, 为黄连产地区分提供研究思路 and 基础。

关键词: 黄连饮片; 支持向量机; 电子鼻; 电子舌; 产地区分; 形状; 气味; 味道; 水分; 总灰分; 醇浸出物; 指标成分; 表小檗碱; 黄连碱; 巴马汀; 小檗碱

中图分类号: R282.71

文献标志码: A

文章编号: 0253-2670(2015)21-3173-06

DOI: 10.7501/j.issn.0253-2670.2015.21.008

Discrimination of *Cotidis Rhizoma* from different habitats by support vector machine

TAO Meng-lin¹, GU Wen-tao¹, WANG Zi-qing¹, HOU Ke-hui¹, CUI Shu-sheng¹, TANG Dao-chao¹, QIN Na¹, ZHANG Da-yong^{2,3}, WAN Jun¹

1. College of Life Science and Engineering, College of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, China

2. Sichuan Neautus Chinese Herbal Slices Co., Ltd., Chengdu 610036, China

3. Key Research Laboratory of Traditional Chinese Medicine Processing Technology, State Administration of Traditional Chinese Medicine, Chengdu 610036, China

Abstract: Objective To establish the identification model based on support vector machine for distinguishing the origin of *Cotidis Rhizoma*. **Methods** Collection of *Cotidis Rhizoma* samples from four habitats, quantification of the external characteristics (including shape, smell, taste, etc.), and according to the current Pharmacopoeia method, determination of characteristic internal values (including moisture, ash, alcohol extract, index composition epiberberine, coptisine, palmatine, berberine content, etc.); In the Matlab 7.0 platform data dimensionality reduction and fusion, to create the habitats for *Cotidis Rhizoma* for the discrimination among the models. **Results** The separate analysis of each datum can not distinguish the habitats of *Cotidis Rhizoma*, while after the use of support vector machine modeling, all superposition feature recognition rate reached 97.1%, which can accurately distinguish the habitats of *Cotidis Rhizoma*; The internal and external feature of high recognition rate is a certain complementarity between each feature subset, the difference of *Cotidis Rhizoma* from different habitats could be comprehensively identified. **Conclusion** The identification model based on support vector machine and the distinction of producing habitats could provide the research ideas and basis for the distinction among the habitats of *Cotidis Rhizoma*.

Key words: *Cotidis Rhizoma*; support vector machine; electronic nose; electronic tongue; habitat discrimination; shape; smell; taste; moisture; total ash content; alcohol extract; index ingredients; epiberberine; coptisine; palmatine; berberine

收稿日期: 2015-06-25

基金项目: 国家“十二五”科技支撑计划项目(2012BAI29B11); 全国大学生科研创新训练计划项目(201510613064)

作者简介: 陶梦琳(1991—), 在读制药工程本科生。Tel: 13518203525 E-mail: tml0702@foxmail.com

*通信作者 万军(1980—), 副教授, 博士, 主要从事中药制剂与炮制研究工作。Tel: 13980561562 E-mail: wwangyidi@126.com

黄连 *Coptidis Rhizoma* 为毛茛科植物黄连 *Coptis chinensis* Franch.、三角叶黄连 *C. deltoidea* C. Y. Cheng et Hsiao 或云连 *C. teeta* Wall. 的干燥根茎^[1]。黄连主产于四川、云南等地,道地性区分是其质量评价的重要内容之一,可从源头保证临床疗效。目前药材产地区分主要依靠人工经验和指纹图谱技术^[2-6],前者主观差异大,准确率较低;后者操作复杂,成本较高。近年来电子鼻和电子舌越来越多地应用到中药气味评价方面^[7-9],其检测快速无损,灵敏度高。支持向量机(SVM)是机器学习领域的研究热点,适用于有限的药材样本,立足“结构风险最小化”原理,较神经网络有更好的泛化能力和理论基础^[6,10-12]。本实验在现有产地区分方法的基础上,依照《中国药典》2010年版标准,对黄连饮片进行相关成分测定,同时借助电子鼻、电子舌采集其气味数据信息;采用支持向量机建立识别模型,实现产地的区分,为黄连产地区分提供研究思路 and 基础。

1 仪器与材料

FOX4000 电子鼻-气味指纹分析仪,含自动进样器 HS100 和 α SOFTV12.3 软件,法国阿尔法莫斯仪器公司;Astree II 电子舌-味觉指纹分析仪、Alpha

M.O.S. α -ASTREE 系统、Alpha M.O.S.自动进样器 LS16、Alpha M.O.S. ASTREE II 软件,法国阿尔法莫斯仪器公司。Agilent 1200 高效液相色谱仪(含四元泵、DAD 检测器、柱温箱、自动进样器)、Chemstation 色谱工作站,美国安捷伦科技有限公司;Welchrom C₁₈ 色谱柱(250 mm×4.6 mm, 5 μ m);Diamonsil(钻石二代)色谱柱(250 mm×4.6 mm, 5 μ m)。游标卡尺,量程 0~150 mm,精度 0.02 mm,成都成量工具有限公司;BS200S 型精密电子天平, $d=1$ mg,最大称量 200 g,北京赛多利斯天平有限公司;FW135 型中草药粉碎机,天津市泰斯特仪器有限公司。Matlab 7.0 软件等。

盐酸、甲醇等试剂均为分析纯。盐酸小檗碱,批号 MUST-13112110,质量分数 $\geq 98\%$,成都曼斯特生物科技有限公司。

黄连药材及饮片:本实验共收集黄连药材 7 批,27 个样本,共加工成饮片 27 份;收集饮片 1 批,4 个样本。药材和饮片均经西南交通大学宋良科副教授鉴定为毛茛科植物黄连 *Coptis chinensis* Franch. (产地为四川、重庆和湖北)和云连 *C. teeta* Wall. (产地为云南)的干燥根茎,详细信息见表 1。

表 1 黄连样品信息

Table 1 Sample information of *Coptidis Rhizoma*

药材产地	采集时间	饮片编号	备注
四川大邑	2014-05	SC1、SC2、SC3	SC1~SC10、CQ1~CQ10、HB1~HB7 由四川新荷花中药饮片股份有限公司 代加工为黄连饮片
四川彭州	2014-05	SC4、SC5、SC6、SC7	
四川洪雅	2014-05	SC8、SC9、SC10	
重庆石柱	2014-05	CQ1、CQ2、CQ3	
重庆石柱	2014-05	CQ4、CQ5、CQ6、CQ7	
重庆石柱	2014-05	CQ8、CQ9、CQ10	直接产地购买饮片
湖北利川	2014-05	HB1、HB2、HB3、HB4、HB5、HB6、HB7	
云南	2014-05	YN1、YN2、YN3、YN4	

2 方法与结果

2.1 内部特征值提取

2.1.1 水分、总灰分、醇浸出物测定 参照《中国药典》2010年版一部黄连饮片水分、总灰分和醇浸出物项下及附录测定方法进行检验^[1]。每个样品测定 3 次,计算平均值。

(1) 水分测定方法:取供试品 2~5 g,平铺于干燥至恒定质量的扁形称量瓶中,厚度不超过 5 mm;疏松供试品不超过 10 mm;精密称定,打开瓶盖在 100~105 $^{\circ}$ C 干燥 5 h,将瓶盖盖好,移置干燥器中,冷却 30 min,精密称定,再在上述温度干

燥 1 h,冷却,称定质量,至连续 2 次称定质量的差异不超过 5 mg 为止。根据减失的质量,计算供试品中含水量。结果见表 2。

(2) 总灰分测定方法:取供试品,粉碎,使能通过 2 号筛,混合均匀后,取供试品 2~3 g,置炽灼至恒定质量的坩埚中,称定质量(准确至 0.01 g),缓缓炽热,注意避免燃烧,至完全炭化时,逐渐升高温度至 500~600 $^{\circ}$ C,使完全灰化并至恒定质量。根据残渣质量,计算供试品中总灰分的质量分数。结果见表 2。

(3) 醇溶性浸出物测定方法:取供试品 2~4 g,

表2 不同产地黄连水分、总灰分及醇浸出物检测结果 ($n=3$)Table 2 Determination of moisture content, total ashes, and extracts in *Cotidis Rhizoma* from each habitat ($n=3$)

编号	水分/%	总灰分/%	醇浸出物/%	编号	水分/%	总灰分/%	醇浸出物/%	编号	水分/%	总灰分/%	醇浸出物/%
SC1	9.84	2.58	29.10	CQ2	8.73	2.29	28.84	HB3	9.63	2.37	32.46
SC2	10.24	2.60	28.64	CQ3	8.93	2.31	29.92	HB4	9.48	2.43	30.99
SC3	10.44	2.50	29.99	CQ4	9.03	2.33	28.73	HB5	9.40	1.95	27.84
SC4	7.19	2.75	31.98	CQ5	9.62	2.17	27.83	HB6	9.22	1.94	29.65
SC5	7.47	2.80	31.18	CQ6	9.22	2.12	29.74	HB7	9.60	1.98	28.76
SC6	7.27	2.89	33.83	CQ7	9.53	2.14	28.05	YN1	9.38	2.36	27.53
SC7	7.53	2.84	32.65	CQ8	8.79	2.20	29.83	YN2	9.63	2.44	27.94
SC8	10.17	2.50	25.06	CQ9	8.94	2.34	31.96	YN3	9.46	2.39	29.39
SC9	10.48	2.45	24.11	CQ10	9.14	2.25	30.17	YN4	9.50	2.35	28.73
SC10	10.49	2.50	24.73	HB1	9.54	2.37	33.28				
CQ1	8.46	2.20	29.39	HB2	9.36	2.36	32.04				

精密称定,置 100~250 mL 的锥形瓶中,精密加乙醇 50~100 mL,密塞,称定质量,静置 1 h 后,连接回流冷凝管,加热至沸腾,并保持微沸 1 h。放冷后,取下锥形瓶,密塞,再称定质量,用乙醇补足减失的质量,摇匀,用干燥滤器滤过,精密量取滤液 25 mL,置已干燥至恒定质量的蒸发皿中,在水浴上蒸干后于 105 °C 干燥 3 h,置干燥器中冷却 30 min,迅速精密称定质量。除另有规定外,以干燥品计算供试品中醇溶性浸出物的质量分数。测定结果见表 2。

由表 2 可见,四川(SC1~SC10)黄连饮片中水分在 7.19%~10.49%,总灰分在 2.45%~2.89%,醇浸出物在 24.11%~33.83%;重庆(CQ1~CQ10)黄连饮片中水分在 8.46%~9.62%,总灰分在 2.12%~2.34%,醇浸出物在 27.83%~31.96%;湖北(HB1~HB7)黄连饮片中水分在 9.22%~9.63%,总灰分在 1.94%~2.43%,醇浸出物在 27.84%~33.28%;云南(YN1~YN4)黄连饮片水分在 9.38%~9.63%,灰分在 2.35%~2.44%,醇浸出物在 27.53%~29.39%。各产地黄连饮片水分、总灰分、醇浸出物测定数据相互重叠,直观分析以上数据未能较好区分各产地黄连饮片。

2.1.2 指标成分测定 参照《中国药典》2010 年版一部黄连饮片项下,采用 HPLC 法对各产地黄连的生物碱类成分进行测定。每个样品测定 3 次,计算平均值。

(1) 对照品溶液的制备:取盐酸小檗碱对照品适量,精密称定,加甲醇制成含盐酸小檗碱 90 $\mu\text{g/mL}$ 的溶液,即得。

(2) 供试品溶液的制备:取各样品粉末(过 2 号筛)约 0.2 g,精密称定,置具塞锥形瓶中,精密加入甲醇-盐酸(100:1)的混合溶液 50 mL,密塞,称定质量,超声处理(功率 250 W,频率 40 kHz) 30 min,放冷,再称定质量,用甲醇补足减失的质量,摇匀,滤过,精密量取续滤液 2 mL,置 10 mL 量瓶中,加甲醇至刻度,摇匀,滤过,取续滤液,即得。

(3) 线性关系考察:精密吸取盐酸小檗碱对照品溶液($\rho=90.4 \mu\text{g/mL}$) 2、4、6、8、10、20 μL 注入液相色谱仪,按《中国药典》2010 年版一部黄连饮片项下的色谱条件测定。以峰面积积分值(A)为纵坐标,以盐酸小檗碱进样量(C)为横坐标绘制标准曲线,得回归方程 $A=2407.5C+16.124$, $r=0.9998$,盐酸小檗碱进样量在 0.18~1.80 μg 内呈良好线性关系。

(4) 测定法:分别精密吸取对照品溶液与供试品溶液各 10 μL ,注入液相色谱仪,测定。以盐酸小檗碱对照品的峰面积为对照,分别计算小檗碱、表小檗碱、黄连碱和巴马汀的量;用待测成分色谱峰与盐酸小檗碱色谱峰的相对保留时间来确定各成分具体峰位。结果见表 3。可见四川(SC1~SC10)黄连饮片中表小檗碱在 0.846 2%~0.995 3%,黄连碱在 1.658 3%~1.802 7%,巴马汀在 1.473 7%~1.685 8%,小檗碱在 5.602 5%~6.479 8%;重庆(CQ1~CQ10)黄连饮片中表小檗碱在 0.807 2%~0.880 2%,黄连碱在 1.341 5%~1.578 3%,巴马汀在 1.468 2%~1.542 9%,小檗碱在 5.682 4%~5.978 5%;湖北(HB1~HB7)黄连饮片中表小檗

表3 不同产地黄连生物碱类成分检测结果 ($n=3$)Table 3 Determination of alkaloids in *Cotidis Rhizoma* from each habitat ($n=3$)

编号	表小檗碱/%	黄连碱/%	巴马汀/%	小檗碱/%	编号	表小檗碱/%	黄连碱/%	巴马汀/%	小檗碱/%
SC1	0.967 3	1.658 3	1.567 2	5.996 8	CQ7	0.859 5	1.578 3	1.524 7	5.726 3
SC2	0.995 3	1.719 1	1.605 3	6.142 8	CQ8	0.842 7	1.341 5	1.489 5	5.732 9
SC3	0.969 3	1.705 8	1.685 8	6.093 2	CQ9	0.827 3	1.352 8	1.507 3	5.683 2
SC4	0.894 3	1.754 7	1.595 8	6.479 8	CQ10	0.807 2	1.375 2	1.542 9	5.803 4
SC5	0.886 5	1.675 5	1.536 7	6.376 4	HB1	0.853 9	1.533 8	1.396 7	5.958 7
SC6	0.877 9	1.693 8	1.600 4	6.196 9	HB2	0.863 4	1.557 5	1.389 4	6.183 7
SC7	0.904 5	1.728 9	1.584 2	6.418 0	HB3	0.875 3	1.557 2	1.455 7	6.072 9
SC8	0.846 2	1.764 4	1.473 7	5.602 5	HB4	0.884 9	1.499 3	1.426 8	5.903 8
SC9	0.852 9	1.802 7	1.538 2	5.832 9	HB5	0.879 3	1.524 8	1.387 9	5.876 1
SC10	0.874 5	1.780 3	1.499 3	5.936 1	HB6	0.865 6	1.563 2	1.395 6	5.995 2
CQ1	0.844 2	1.505 3	1.468 2	5.978 5	HB7	0.861 7	1.534 2	1.436 5	5.803 8
CQ2	0.873 9	1.563 6	1.495 7	5.792 4	YN1	0.824 3	1.376 5	1.436 1	5.810 4
CQ3	0.880 2	1.558 2	1.501 7	5.736 6	YN2	0.817 8	1.388 9	1.486 2	5.892 6
CQ4	0.847 1	1.498 4	1.523 3	5.873 1	YN3	0.831 7	1.457 2	1.458 2	5.903 1
CQ5	0.852 1	1.505 8	1.504 2	5.682 4	YN4	0.824 3	1.376 5	1.436 1	5.810 4
CQ6	0.876 4	1.523 9	1.483 9	5.923 6					

碱在 0.853 9%~0.884 9%，黄连碱在 1.499 3%~1.563 2%，巴马汀在 1.387 9%~1.455 7%，小檗碱在 5.803 8%~6.183 7%；云南（YN1~YN4）黄连饮片中药小檗碱在 0.817 8%~0.831 7%，黄连碱在 1.376 5%~1.457 2%，巴马汀在 1.436 1%~1.486 2%，小檗碱在 5.810 4%~5.903 1%。各产地黄连饮片中生物碱类成分测定数据相互重叠，直观分析以上数据仍未能较好区分各产地黄连饮片。

2.2 外部特征值提取

2.2.1 外观性状测定 通过化学成分等内在指标检测，不能较好区分各黄连饮片产地。在以上实验基础上，进一步采集性状大小、气味、味道等外部数

据信息。测定各黄连饮片长度、宽度及厚度，每个样品取 50 份，测定后计算平均值。结果见表 4。四川（SC1~SC10）黄连长度在 22.95~53.83 mm，宽度在 4.05~8.52 mm，厚度在 2.76~3.93 mm；重庆（CQ1~CQ10）黄连长度在 18.58~33.16 mm，宽度在 4.92~7.17 mm，厚度在 1.96~3.61 mm；湖北（HB1~HB7）黄连长度在 14.97~29.54 mm，宽度在 4.86~7.64 mm，厚度在 2.78~4.82 mm；云南（YN1~YN4）黄连长度在 25.61~42.88 mm，宽度在 5.36~8.33 mm，厚度在 2.17~3.44 mm。各产地黄连饮片外观测定数据相互重叠，直观分析以上数据仍未能较好区分各产地黄连饮片。

表4 不同产地黄连饮片长度、宽度及厚度检测结果 ($n=50$)Table 4 Length, width, and thickness of *Cotidis Rhizoma* pieces from each habitat ($n=50$)

编号	长度/mm	宽度/mm	厚度/mm	编号	长度/mm	宽度/mm	厚度/mm	编号	长度/mm	宽度/mm	厚度/mm
SC1	42.11	7.46	3.53	CQ2	27.52	5.75	2.72	HB3	17.68	5.02	2.78
SC2	53.83	8.03	3.60	CQ3	31.07	5.91	2.81	HB4	21.07	6.25	4.82
SC3	24.98	4.94	2.87	CQ4	18.85	6.76	1.96	HB5	25.12	4.99	4.27
SC4	22.95	4.05	2.86	CQ5	18.58	5.14	2.61	HB6	19.85	5.06	3.85
SC5	27.54	6.24	2.76	CQ6	29.04	7.17	3.19	HB7	14.97	4.86	2.99
SC6	31.56	6.92	3.17	CQ7	14.25	4.92	2.03	YN1	25.61	6.16	2.75
SC7	36.01	7.63	3.21	CQ8	33.16	6.67	3.61	YN2	42.88	8.33	3.44
SC8	28.85	6.07	3.07	CQ9	25.54	5.02	3.42	YN3	37.25	7.84	3.12
SC9	47.58	8.52	3.93	CQ10	23.78	6.43	3.01	YN4	26.82	5.36	2.17
SC10	37.91	7.55	3.24	HB1	27.62	7.64	3.55				
CQ1	20.31	7.03	2.14	HB2	29.54	6.95	4.23				

2.2.2 气味性状测定

(1) 电子鼻采集气味数据: 取生黄连, 打粉过3号筛。精密称取黄连粉末1.0 g置于20 mL顶空进样瓶中, 顶空进样, 平行测定3次^[7,13]。电子鼻共包含有18个金属传感器, 分别为LY2/LG、LY2/G、LY2/AA、LY2/GH、LY2/gCTL、LY2/gCT、T30/1、P10/1、P10/2、P40/1、P70/2、PA/2、P30/1、P40/2、P30/2、T40/2、T40/1、TA/2, 因此每个样品有18组数据(每1根传感器的数据作为1维特征值, 共有18维特征值)。数据采集120 s。实验以120 s内传感器最大特征响应值作为输出值。各产地黄连

LY2/LG传感器原始数据见表5。其余传感器原始数据与此类似, 此处不再赘述。

(2) 电子舌采集味道数据: 取生黄连, 粉碎, 过3号筛, 精密称定粉末1.0 g于250 mL锥形瓶中, 加水80 mL, 浸泡30 min, 加热回流1 h, 放冷, 滤过, 取滤液20 mL, 加水定容至100 mL, 备用^[14-15]。电子舌包含7个脂质膜传感器, 分别为ZZ、AB、GA、BB、CA、DA、JE, 因此每个样品有7组数据(每1根传感器的数据作为1维特征值, 共有7维特征值)。各产地黄连ZZ传感器原始数据见表5。其余传感器原始数据与此类似, 此处不再赘述。

表5 各产地黄连电子鼻(LY2/LG)和电子舌(ZZ)原始数据

Table 5 Primary data of *Cotidis Rhizoma* from each habitat by EN (LY2/LG) and ET (ZZ)

编号	LY2/LG	ZZ	编号	LY2/LG	ZZ
SC1	0.033 383 158 943 697 3	1 875.028 371 428 57	CQ7	0.037 279 596 977 331 2	1 836.019 157 142 86
SC2	0.035 696 073 431 922 3	1 873.444 757 142 86	CQ8	0.063 668 519 454 271 3	1 827.374 657 142 86
SC3	0.037 678 207 739 308 7	1 872.093 600 000 00	CQ9	0.069 437 404 967 055 1	1 824.120 257 142 86
SC4	0.038 384 845 463 610 3	1 870.437 342 857 14	CQ10	0.070 095 814 422 590 8	1 822.623 814 285 71
SC5	0.038 617 886 178 861 7	1 868.185 414 285 71	HB1	0.048 719 236 564 539 7	1 827.505 414 285 71
SC6	0.040 567 951 318 459 7	1 865.003 657 142 86	HB2	0.052 578 361 981 800 5	1 825.137 257 142 86
SC7	0.040 485 829 959 515 2	1 862.940 600 000 00	HB3	0.055 919 395 465 996 8	1 825.398 771 428 57
SC8	0.026 500 000 000 001 6	1 856.199 342 857 14	HB4	0.054 718 875 502 007 4	1 819.587 342 857 14
SC9	0.024 860 476 915 271 9	1 855.632 728 571 43	HB5	0.050 403 225 806 451 1	1 829.190 728 571 43
SC10	0.027 327 935 222 675 1	1 853.787 600 000 00	HB6	0.053 725 291 434 367 3	1 828.101 085 714 29
CQ1	0.030 515 257 628 814 7	1 840.101 685 714 29	HB7	0.056 088 933 804 953 8	1 826.662 757 142 86
CQ2	0.030 364 372 469 638 5	1 838.736 000 000 00	YN1	0.064 974 619 289 340 1	1 841.598 128 571 43
CQ3	0.032 290 615 539 858 7	1 835.757 642 857 14	YN2	0.068 228 105 906 313 6	1 838.300 142 857 14
CQ4	0.033 7531 486 146 094	1 833.142 500 000 00	YN3	0.069 331 983 805 667 9	1 836.106 328 571 43
CQ5	0.037 0741 482 965 946	1 841.263 971 428 57	YN4	0.072 791 164 658 633 2	1 833.098 914 285 71
CQ6	0.036 924 633 282 751 6	1 839.491 485 714 29			

通过采集气味及味道数据, 并采用软件进行主成分分析(PCA)等化学计量方法分析, PC1、PC2和PC3三者总贡献率仅为82.1%(区分度越大越好, 一般大于90%有一定区分度), 区分度不理想。

2.3 基于支持向量机的数据融合与处理

2.3.1 内外部数据的融合 为了保证样本的代表性和实验的客观性, 本实验样本总数为31个。对水分、总灰分、醇溶性浸出物、表小檗碱、黄连碱、巴马汀、小檗碱质量分数7维内部特征数据; 气味的18维外部特征数据, 味道的7维外部特征数据和长度、宽度、厚度3维外部形态特征数据进行融合, 每个黄连样本得到35维特征, 总体的特征矩阵为31×35。特征矩阵经过最小二乘支持向量机(LSSVM)

的学习、训练和识别, 最终输出为各个黄连样本的归属地。本研究中使用Toolbox: LSSVMlabv1_8软件进行数据处理。

2.3.2 数据归一化预处理 为了消除不同属性数据的量纲的不同, 对每个属性的值进行了归一化, 将属性值映射到了[-1, 1]。归一化的表达式为 $y = (y_{\max} - y_{\min}) \times (x - x_{\min}) / (x_{\max} - x_{\min}) + y_{\min}$, 其中 x 为样本的属性值, y 表示归一化后的上下确界。

2.3.3 实验结果 从数据集中, 随机选取了每类样本中的50%作为训练集, 剩余的50%作为测试集。实验共重复了100次。仿真得到的结果见表6。可以看出, 不同的特征子集的识别率表现出较大差异。其中, 电子舌和内部特征识别率较高, 平均可以达

表6 不同特征子集的认可率对比 (括号内数据为标准差)

Table 6 Recognition rates with different characteristic subset (numbers in brackets are standard values)

分类器	核函数	识别率/%				
		电子鼻属性	电子舌属性	统计内部数据	统计外部数据	全部属性
LSSVM	linear	62.2 (11.9%)	93.2 (9.8%)	91.6 (10.1%)	81.7 (9.4%)	97.1 (10.1%)
	nonlinear	66.1 (13.6%)	92.6 (7.9%)	94.4 (7.6%)	77.7 (10.1%)	95.6 (7.5%)

到90%以上,表明这2种特征类型更能反映类间差异,可作为有效区分黄连饮片产地的模型。

3 讨论

现在对产地区分往往针对中药材,对饮片的产地辨识较少。这也出现一些非道地产区药材运往道地产区进行饮片炮制,然后贴牌按道地药材(饮片)销售获取利润的现象。而饮片产地区分多和饮片鉴别相联系^[16-18],往往靠传统经验,主观性较强。前期研究中,课题组已采用电子鼻成功区分了黄连及其炮制品^[7]。

本研究中对核函数的种类进行了选择,线性核函数的表达式为 $K(x_i, x_j) = x_i^T x_j$,非线性核函数的表达式为 $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ 。通过实验可以看出,针对实验中的数据,使用线性核函数的分类器的平均准确率要高一些。线性核函数是在原始空间中寻找最优泛化性的线性分类器,而非线性核函数如高斯核有可能出现过学习的现象。所以针对不同的数据集,需要选择合适的核函数。

标准差即识别率的波动较大。针对不同的样本得到的结果变化范围较大,说明实验结果与样本的数量有较大的关系。为了减小识别率的波动,需要增加更多的样本,弱化不同训练集之间的差别。

电子舌相对于电子鼻而言,其识别率更高,这也符合黄连以“味苦者佳”的传统经验鉴别。而所有特征叠加使用时能够得到更高的识别率97.1%,说明各特征子集间具有一定的互补性,从不同侧面反映了来自不同产地黄连饮片间的差异性。在味道数据特征和所有叠加特征均能有效区分产地的前提下,基于支持向量机的电子舌鉴别模型更具有操作简便、无损、快速的优点。

参考文献

- [1] 中国药典 [S]. 一部. 2010.
- [2] 肖小河, 陈士林, 黄璐琦, 等. 中国道地药材研究20年概论 [J]. 中国中药杂志, 2009, 34(5): 519-523.
- [3] 秦雪梅, 孔增科, 张丽增, 等. 中药材“辨状论质”解读及商品规格标准研究思路 [J]. 中草药, 2012, 43(11):

2093-2098.

- [4] 李强, 杜思邈, 张忠亮, 等. 中药指纹图谱技术进展及未来发展方向展望 [J]. 中草药, 2013, 44(22): 3095-3014.
- [5] 胡汉昆, 刘薇芝, 刘萍, 等. 中药特征图谱技术在中药鉴定中的应用 [J]. 中国医院药学杂志, 2009, 29(11): 936-938.
- [6] 常建丽. 基于机器嗅觉的中药材质量检测技术的研究 [D]. 金华: 浙江师范大学, 2012.
- [7] 刘瑞新, 李慧玲, 李学林, 等. 基于电子舌的穿心莲水煎液的掩味效果评价研究 [J]. 中草药, 2013, 44(16): 2240-2245.
- [8] 孙玉平, 张铁军, 曹煌, 等. 中药辛味药性表达及在临证配伍中的应用 [J]. 中草药, 2015, 46(6): 785-790.
- [9] 宁珂. 电子鼻与电子舌融合技术及其应用 [D]. 吉林: 东北电力大学, 2014.
- [10] 龙伟, 刘培勋, 曾平, 等. 支持向量机与中药研究 [J]. 中草药, 2008, 39(10): 1441-1445.
- [11] 王国胜. 支持向量机的理论算法研究 [D]. 北京: 北京邮电大学, 2007.
- [12] 吴莎, 刘启安, 李云, 等. 最小二乘支持向量机和偏最小二乘法在栀子中间体纯化工艺近红外定量分析中的应用 [J]. 中草药, 2015, 46(7): 990-997.
- [13] 万军, 周霞, 黄永亮, 等. 天麻配方颗粒制备中气味相关性研究 [J]. 中草药, 2013, 44(7): 825-828.
- [14] Han F, Huang X, Teye E, et al. Nondestructive detection of fish freshness during its preservation by combining electronic nose and electronic tongue techniques in conjunction with chemometric analysis [J]. Anal Methods, 2014, 6(2): 529-536.
- [15] Li Y J, Lei J C, Yang J N, et al. Classification of tieguanyin tea with an electronic and pattern recognition [J]. Anal Lett, 2014, 47(14): 2361-2369.
- [16] 吴娜, 杨诗龙, 严丹, 等. 粉末中药鉴别方法的研究进展与思考实践 [J]. 中草药, 2015, 46(10): 1413-1419.
- [17] 刘淑兰, 曹蕴红, 刘艳, 等. 中药鉴定技术的发展及文献研究分析 [J]. 中医药信息, 2008, 25(2): 78-79.
- [18] 刘静, 周晓梅, 祝与鸣, 等. 中药质量控制方法研究进展 [J]. 中国药房, 2010, 21(3): 281-282.