

基于烟叶化学成分的烟叶间可替代性评价

张其龙, 易浩, 任建新, 黄振军, 杜文

湖南中烟工业有限责任公司, 长沙市劳动中路426号410007

摘要: 研究了基于烟叶化学成分数据的烟叶间可替代性评价方法。首先对每类烟叶进行分布式取样; 然后检测样品的总糖、总氮、总碱、氯、纤维素、淀粉、灰分、多酚、醚提取物总量和总挥发碱等10项主要化学成分含量, 为每类烟叶构造化学成分数据集; 接下来计算未知烟叶数据集中各样本与目标烟叶集的马氏距离, 以平均距离的大小评价未知烟叶替代目标烟叶的可行性。用115类烟叶对方法进行了交叉检验: 每类烟叶取一个样品构成检验集, 计算检验集与去掉检验集样品的115类烟叶间的距离, 结果检验集样品被判断为与其原来所属的类很相似的比率高于80%, 表明评价方法具有较高的可靠性。

关键词: 烟叶; 替代性; 化学成分; 马氏距离

doi:10.3969/j.issn.1004-5708.2010.z1.007

中图分类号: TS411 文献标识码: A 文章编号: 1004-5708(2010)增刊-0028-05

Replaceability evaluation in tobacco leaves based on chemical compositions

ZHANG Qi-long, YI Hao, REN Jian-xin, HUANG Zhen-jun, DU Wen

China Tobacco Hunan Industrial Corp. Ltd., 426 Middle Laodong Road, Changsha 410007, China

Abstract: A method of evaluating tobacco leaf replaceability was developed using chemical composition data. Ten chemical constituents, i.e. sugar, nitrogen, alkaline, chloride, cellulose, ash, starch, polyphenols, petrol-ether extract and volatile base of tobacco were determined. The Mahalalobis distances of unknown tobacco samples and data set of destination tobacco was calculated. The mean distance represented replaceability of unknown tobacco to destination tobacco. The distances of 115 independent validation samples from 115 classes of destination tobacco of different district, grade and variety were calculated to validate the method. More than 80% validation samples were classified correctly.

Key words: tobacco; replaceability; chemical composition; mahalalobis distance

卷烟产品设计人员为了维持产品风格特色的稳定, 需要从不同产地、品种、等级的烟叶中选择原料, 组合成叶组配方, 由于每种烟叶原料的数量是有限的, 因此叶组配方不得不进行频繁的调整, 在某种烟叶原料用尽后寻找其替代烟叶。传统上, 只能依赖繁重的单料烟评吸以及试制烟制作来确定合适的替代叶组配方。另外, 在配方打叶设计中, 由于打叶复烤厂地点分散, 对配打烟叶的评吸更难组织, 配打叶组配方很大程

度上依赖经验和以往对烟叶的印象, 存在较大的风险。因此, 若能通过烟叶的客观检测数据来评价不同烟叶间的可替代性, 对于卷烟产品设计和配方打叶设计都具有十分重要的意义^[1]。目前, 烟草加工企业普遍应用烟叶常规化学成分来作叶组配方时的参考^[2], 然而糖、氮、碱、氯等常规化学成分仅能代表烟叶整体特征的少部分, 在烟叶的内在质量评价中能起到的作用还非常有限^[3], 简单的常规化学成分对比不足以支撑烟叶的可替代性评价工作。本文研究了用更加全面的10项烟叶化学成分数据: 总糖、总氮、总碱、氯、纤维素、淀粉、多酚、灰分、醚提取物总量和总挥发碱来评价烟叶间可替代性的方法。

1 方法

1.1 材料

2004年度片烟按产地、等级和品种分为115类, 每

作者简介: 张其龙, 男, 工程师, 主要从事烟叶质量评价和打叶复烤工艺研究。

杜文(通讯作者), 男, 副研究员, 主要从事烟草化学方面的研究工作, Tel: 0731-85559111-22112, E-mail: duwen@baisha.com

基金项目: 中国烟草总公司基金资助项目: 部分替代进口烟叶生产及工业验证(国烟科[2006]371号)

收稿日期: 2010-05-18

类烟叶总量 500 ~ 10000 件(200 kg/件),对每类烟叶分布式取样 12 ~ 50 个,“分布式取样”意为使取样点尽量均匀地分布在总体母本中,如在某类烟叶 2000 件片烟中每 100 件取样一个,可取样 20 个。在 115 类烟叶中共取样 3104 个,每个样品约 100g。样品 40℃ 烘干后经旋风磨粉碎,过 40 目筛。

外部检验样品 115 个:由 115 类烟叶中每类各取 1 个样品构成,分别编为 T1 ~ T115。

目标烟叶样品集 115 个:由每类烟叶样品集中去掉外部检验样品后所剩样品构成,分别编为 C1 ~ C115,共有 2989 个样品。

编号序号相同的外部检验样品和目标烟叶集样品具有相同的产地、等级和品种属性。

1.2 仪器

MPA 型近红外光谱仪(德国 Bruker 公司),配积分球和旋转样品台采样附件;Cyclotec 1093 型旋风磨(瑞典 Foss 公司)。

1.3 样品近红外光谱扫描

近红外光谱扫描条件:采集样品漫反射近红外光谱,分辨率 8 cm^{-1} ,光谱范围 $4000\text{ cm}^{-1} \sim 9000\text{ cm}^{-1}$,重复扫描 64 次。

1.4 样品化学成分检测

用烟草化学成分近红外分析模型预测样品的化学成分含量,共 10 项指标:总糖、总氮、总碱、氯、纤维素、淀粉、多酚、灰分、醚提取物总量和总挥发碱。

1.5 烟叶样品与目标烟叶样品集之间的距离计算

距离计算在 Matlab R2009b 软件平台完成。

设有 n 种目标烟叶 $C_1 \sim C_n$,第 i 种目标烟叶的取样数量为 m_i ,由目标烟叶样品的多种主要化学成分指标的检测值构成了目标烟叶化学成分值矩阵 A_i 。检测未知烟叶与目标烟叶相同化学成分指标,得到未知烟叶化学成分值向量 t 。计算未知烟叶化学成分值向量 t 与目标烟叶 i 化学成分值矩阵 A_i 的马氏距离:

1) 计算目标烟叶 i 化学成分值矩阵 A_i 的均值中心向量 a_i ,对矩阵 A_i 进行均值中心化处理得到中心化矩阵 X_i ;

2) 计算未知烟叶化学成分值向量 t 与目标烟叶化学成分值矩阵的马氏距离:

$$D_i = (t - a_i) \times (X_i' \times X_i)^{-1} \times (t - a_i)' \times (m_i - 1) \quad (\text{式 } 1)$$

重复以上步骤,直至得到未知烟叶与所有 n 种目标烟叶集间的距离 $D_1 \sim D_n$,对距离 $D_1 \sim D_n$ 按从小到大的顺序排序,距离越小表明相似度越高。

1.6 烟叶间的可替代性评价

按“式 1”计算未知烟叶集 C_x 中各样本与目标烟叶集 TC_y 的距离,以距离的平均值大小评价以未知烟叶 C_x 替代目标烟叶 TC_y 的可行性。

2 结果与讨论

2.1 烟叶样品集的构成

单比较两个烟叶样品的相似度可用性不高,首先难以保证取样的代表性,在目前对片烟进行外观质量评价存在困难的情形下更是如此^[4]。另外由于批次烟叶存在质量波动的问题^[5],对未知烟叶的相似性判断应与目标烟叶的分布状态有关。图 1 为两种烟叶的总糖含量多次取样检测结果,从图中可见,产地 1-C3F 与产地 2-C3F 的总糖均值接近,分布状况则大不相同,产地 1-C3F 烟叶总糖 40 次取样检测平均值 20.0%,标准偏差为 4.2%,产地 2-C3F 烟叶总糖 33 次取样检测平均值为 19.1%,标准偏差为 1.3%,按工业可用性的要求,假设只考虑总糖因素,虽然两种烟叶的平均水平非常相似,也不能用产地 1-C3F 替代产地 2-C3F 使用。

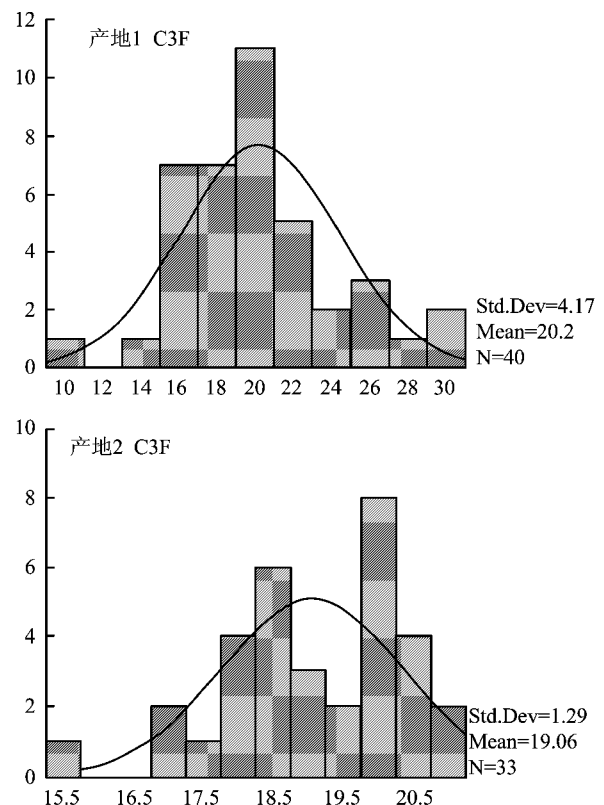


图 1 产地 1 和产地 2C3F 烟叶中总糖多次取样检测的结果
注:图中横轴为样品总糖含量/%,柱高为区间内的样品数,曲线为正态分布拟合曲线

为了能正确地表达目标烟叶的平均水平和波动水平,在建模时需要对每种目标烟叶进行多次分布式取样,构成目标烟叶样品集。本文根据目标烟叶实际批量的大小,每种目标烟叶取样12~50个。

2.2 化学成分检测指标的选择

马氏距离计算要求样本数量大于指标数量,如果选用指标过多会加大取样和检测的压力,根据取样的可行性,选择10个左右化学成分指标较合适。在选择计算所用化学成分指标时,一方面考虑其重要性、代表性和检测效率,另一方面考虑各个指标之间要保持一定的独立性,即在计算中不重复考虑线性相关性较强的指标。图2中262个样品的检测数据线性相关性考察结果表明,总糖-还原糖之间有较强的相关性,总碱-总氮之间有一定的相关性,最终选择总糖、总氮、总碱、氯、纤维素、淀粉、多酚、灰分、醚提取物总量和总挥发碱共10项指标来作烟叶相似性计算。

2.3 距离计算方法的检验

在115类共3104个烟叶样品中,每类中取1个样品得到115个外部检验样品:T1~T115,用其余2989个烟叶样品构成了115个目标烟叶集:C1~C115。检测所有样品的总糖、总氮等10项烟叶化学成分指标,计算每个外部检验样品与每个目标烟叶集的距离,得到115×115距离数据,对方法进行检验。

表1为搜索与目标烟叶集相似的未知烟叶的结果,括号内为检验样品与目标烟叶模型间的距离值D。因篇幅所限,仅列出了与40个目标烟叶集C1~C40距离最近的前7个检验样品。10因素的马氏距离计算

结果服从自由度为10的卡方分布(χ_{10}^2)^[6],可以根据置信度水平以及样品与目标集的距离设定:距离<18的为“高度相似”,距离在18到36间为“很相似”,距离36到54间为“较相似”,距离54到100间为“弱相似”,距离大于100为“不相似”。表1显示,以“很相似”(距离值<36)为正确判断的标准,大部分检验样品被判断为与其原来所属的类很相似,在表1中列出的40个目标烟叶集的相似样品搜索结果中,共7个样品超出此范围,模式识别准确率>80%。

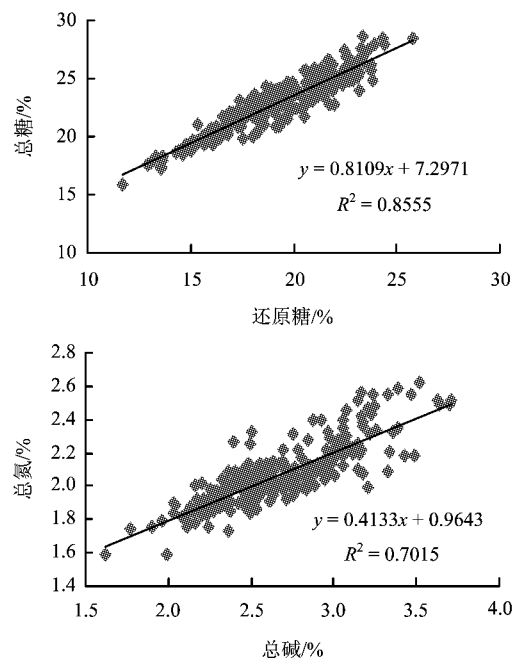


图2 烟叶总糖-还原糖、总氮-总碱的线性相关性

表1 烟叶相似性计算模型外部检验结果——与目标烟叶集最接近的7个外部检验样品

类名	烟叶	1	2	3	4	5	6	7
C1	巴西片 BLO1	T1(18)*	T4(34)	T2(36)	T26(132)	T3(166)	T72(411)	T27(419)
C2	巴西片 BOC-S	T2(8)	T4(40)	T3(100)	T1(193)	T26(257)	T13(372)	T21(374)
C3	巴西片 MDC-S	T3(8)	T2(109)	T48(119)	T27(131)	T4(141)	T99(198)	T56(208)
C4	巴西片 MOA-S	T4(6)	T2(39)	T1(96)	T3(103)	T26(111)	T42(238)	T53(253)
C5	龙山 C3F	T41(18)	T5(27)	T44(41)	T88(107)	T83(155)	T6(181)	T17(194)
C6	龙山 CB	T41(45)	T6(45)	T49(64)	T5(66)	T91(76)	T48(93)	T42(103)
C7	滨县 C2L	T7(7)	T8(71)	T68(122)	T96(224)	T86(412)	T36(552)	T63(585)
C8	滨县 C3L	T8(12)	T7(32)	T68(50)	T86(58)	T96(69)	T80(82)	T36(122)
C9	威丰 C3L	T12(29)	T9(29)	T15(45)	T100(46)	T84(50)	T92(51)	T11(55)
C10	襄樊 C4F	T10(25)	T28(43)	T100(45)	T11(71)	T76(74)	T50(78)	T17(85)
C11	宣恩 C3F	T11(20)	T12(27)	T9(44)	T115(54)	T104(59)	T10(66)	T17(81)
C12	宣恩 C3L	T115(31)	T77(32)	T11(36)	T100(39)	T95(59)	T83(62)	T99(70)
C13	桂阳 B2F	T13(96)	T52(136)	T53(143)	T59(203)	T46(439)	T41(477)	T42(487)
C14	桂阳 C2F	T14(30)	T15(38)	T84(42)	T38(56)	T40(58)	T28(59)	T16(60)
C15	桂阳 C3F	T15(12)	T16(27)	T14(27)	T84(37)	T92(44)	T39(47)	T48(52)
C16	桂阳 CX-2	T16(21)	T92(31)	T84(43)	T15(49)	T63(52)	T14(60)	T51(67)

(续表 1)

C17	桑植 C3F	T17(38)	T88(41)	T91(63)	T83(64)	T28(77)	T61(79)	T106(84)
C18	桑植 B1F	T44(83)	T20(97)	T18(99)	T21(169)	T17(257)	T5(269)	T91(282)
C19	桑植 B1L	T41(46)	T19(81)	T20(121)	T91(145)	T83(172)	T40(185)	T15(190)
C20	桑植 B2F	T19(14)	T81(20)	T20(23)	T18(28)	T106(70)	T44(72)	T53(76)
C21	桑植 B2L	T21(33)	T19(38)	T17(47)	T106(54)	T78(59)	T44(68)	T81(68)
C22	桑植 C3L	T83(21)	T22(23)	T91(34)	T17(34)	T51(41)	T15(44)	T99(48)
C23	桑植 B4	T23(29)	T53(124)	T52(130)	T59(141)	T18(151)	T54(155)	T89(172)
C24	永州 C3L	T24(17)	T51(37)	T55(82)	T49(89)	T15(104)	T56(126)	T22(141)
C25	大安 C3L	T25(7)	T68(385)	T80(423)	T86(458)	T7(513)	T96(549)	T8(609)
C26	津巴布韦 BIOF	T26(7)	T1(48)	T72(57)	T4(102)	T2(107)	T3(114)	T48(125)
C27	津巴布韦 C2O	T55(24)	T27(35)	T80(68)	T48(78)	T3(96)	T24(100)	T26(114)
C28	古蔺 C3FA	T28(20)	T15(50)	T16(60)	T115(72)	T76(72)	T84(80)	T77(80)
C29	攀枝花 C3F	T29(34)	T93(88)	T77(105)	T82(121)	T104(122)	T11(149)	T61(174)
C30	楚雄 C2F	T30(20)	T31(22)	T33(29)	T34(58)	T76(73)	T29(77)	T32(90)
C31	楚雄 C3F	T31(18)	T30(31)	T34(59)	T104(64)	T77(77)	T33(84)	T85(114)
C32	昆明 C3F	T32(100)	T85(147)	T35(220)	T34(280)	T33(287)	T77(341)	T30(373)
C33	弥勒 C4F	T34(29)	T30(41)	T33(53)	T36(55)	T29(61)	T77(71)	T93(99)
C34	师宗 C3FA	T34(17)	T30(30)	T35(45)	T33(62)	T77(66)	T32(69)	T29(84)
C35	师宗 C4F	T35(16)	T32(89)	T85(105)	T34(168)	T80(180)	T33(233)	T30(274)
C36	泸西 X2F	T36(5)	T105(142)	T108(187)	T86(213)	T76(272)	T39(322)	T107(347)
C37	衡南 B2F	T37(17)	T59(54)	T52(111)	T13(133)	T66(151)	T53(151)	T6(161)
C38	衡南 C2F	T38(27)	T15(29)	T39(33)	T16(42)	T92(44)	T84(49)	T37(77)
C39	衡南 C3F	T39(21)	T15(28)	T16(30)	T51(39)	T14(44)	T28(65)	T84(66)
C40	江华 C3FA	T40(29)	T74(145)	T51(147)	T62(153)	T5(179)	T63(205)	T6(220)

注: * 括号内的数值为外部检验样品与目标烟叶集的距离计算结果。

表 2 列出了与表 1 相反方向的计算结果——搜索与外部检验样品最接近的 7 个目标烟叶集,因篇幅所限仅列出 10 个外部检验样品的最相近的 7 个搜索结果。

果。表 2 数据显示,10 个检验样品的相似目标烟叶集搜索结果中有 9 个判断正确,表明本文所用距离计算方法具有较高的准确性。

表 2 相似烟叶搜索结果——与外部检验样品最相似的 7 个目标烟叶集

外部检验样品	1	2	3	4	5	6	7
T1	C1(18)*	C26(48)	C42(70)	C4(96)	C48(126)	C54(151)	C27(152)
T2	C2(8)	C1(36)	C4(39)	C26(107)	C3(109)	C27(117)	C48(132)
T3	C3(8)	C48(92)	C27(96)	C15(100)	C2(100)	C4(103)	C26(114)
T4	C4(6)	C1(34)	C2(40)	C26(102)	C27(138)	C3(141)	C54(162)
T5	C44(18)	C5(27)	C46(40)	C6(66)	C58(92)	C59(100)	C42(103)
T6	C6(45)	C46(57)	C58(69)	C44(69)	C60(74)	C62(92)	C66(106)
T7	C7(7)	C8(32)	C68(66)	C109(156)	C96(195)	C80(209)	C60(228)
T8	C8(12)	C7(71)	C68(72)	C96(93)	C109(113)	C106(146)	C80(146)
T9	C9(29)	C11(44)	C50(56)	C62(57)	C12(72)	C39(86)	C22(89)
T10	C10(25)	C106(59)	C11(66)	C104(79)	C12(87)	C102(100)	C9(101)

注: * 括号内的数值为外部检验样品与目标烟叶集的距离计算结果。

2.4 烟叶间的可替代性评价

计算烟叶数据集 C_x 中各样本与目标烟叶数据集 TC_y 的距离,以距离的平均值为替代性评价指数,衡量

以烟叶 C_x 替代烟叶 TC_y 的可行性。依此计算 115 类烟叶间的替代性评价指数,结果见表 3,因篇幅所限,仅列出了前 15 类烟叶间的可替代性关系。可以参照

“2.3”中的方法设定“高”、“中”、“低”、“弱”、“无”等替代可行性的数值范围为： < 18 、 $18 \sim 36$ 、 $36 \sim 54$ 、 $54 \sim 100$ 、 > 100 。值得注意的是，由于替代性评价指数与烟叶的分布状态相关，烟叶 C_x 对 TC_y 的替代性评价指数与烟叶 C_y 对 TC_x 的替代性评价指数值往往是不同的，

这与烟草工业企业现实中对烟叶替代使用的认识是一致的：对于两种中心水平接近的烟叶，用一致性较高的烟叶替代一致性较差的烟叶具有较小的风险，反之则有较大的风险。

表3 烟叶 C_x 对目标烟叶 TC_y 的替代性评价指数

类名	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	TC11	TC12	TC13	TC14	TC15
C1	—	71	139	46	1785	332	5598	1019	483	516	312	471	389	357	254
C2	26	—	89	25	2233	365	5117	1023	456	527	260	459	463	336	225
C3	130	114	—	95	1082	374	3334	708	217	472	243	286	673	155	101
C4	35	49	149	—	2145	341	6583	1248	619	657	329	592	468	406	281
C5	1011	1038	886	679	—	61	6693	1073	548	1183	458	276	445	240	281
C6	1101	989	559	736	185	—	4118	692	208	517	225	153	821	163	174
C7	1176	1434	1028	1331	849	566	—	29	967	541	507	265	8675	296	240
C8	1237	1645	822	1486	875	488	99	—	837	493	469	197	7682	275	236
C9	891	723	340	744	462	178	1819	366	—	69	55	33	1507	140	111
C10	848	668	271	708	587	233	2038	402	71	—	55	48	1571	182	190
C11	667	621	392	722	590	365	2016	339	71	81	—	47	1843	191	230
C12	968	765	407	817	550	327	1529	307	46	88	34	—	2483	178	195
C13	631	554	594	433	698	633	7226	1147	522	1525	586	556	—	197	110
C14	1036	923	592	715	388	256	2416	394	120	668	214	183	1834	—	28
C15	1120	931	496	774	396	204	2354	393	90	480	175	133	1482	27	—

3 结论

115类烟叶之间的交叉检验结果表明，检验样品能被正确地判断出与其原来所属的类最相似，表明方法在评价烟叶相似性方面是有效的，烟叶10个指标集的接近程度可以较好地代表烟叶整体的相似程度，基于该方法的烟叶可替代性评价结果可为产品设计人员选择替代烟叶提供有益的参考。此外，烟叶间的可替代性计算结果(表3)表明，两类烟叶间的可替代性可能是不对称的，这符合产品设计对一致性水平不同的烟叶的工业可用性的认识。

由于烟叶化学成分种类繁多，我们能方便获得的化学成分数据所包含的信息不能充分表达烟叶的整体化学构成，因此在距离计算中必然存在缺失之处，影响了用化学成分进行烟叶模式识别的准确率。虽然本文仅选用了10项烟叶常量化学成分，但因为烟叶是自然生长的农作物，主要常量化学成分的状态一定程度上代表了烟草的整体状态，因此用于对烟叶整体相似性进行判断时也有较高的准确性，已具有一定的实用价值。随着更多化学成分指标的加入，化学成分对烟草

整体状态的表达会增强，烟叶模式识别的可靠性会更高。而且与基于指纹图谱等模糊信息的模式识别方法^[7]比较，用样品化学成分进行模式识别计算可以明确了解导致分类差异的因素，在应用上具有一定优势。

参考文献

- [1] Dale Hill. 迎接挑战——认识烟叶的质量和可用性[J]. 烟草科技, 1997(1):34-36.
- [2] 薛超群, 尹启生, 王信民, 等. 烤烟烟叶香气质量与其常规化学成分的相关性[J]. 烟草科技, 2000(9):27-30.
- [3] 朱尊权. 烟叶的可用性与卷烟的安全性[J]. 烟草科技, 2000(8):3-6.
- [4] 杨明峰. 片烟外观质量评价模式探讨[J]. 中国烟草科学, 2007, 28(4):18-21.
- [5] 杜文, 谭新良, 易建华, 等. 用烟叶化学成分进行烟叶质量评价[J]. 中国烟草学报, 2007, 13(3):25-31.
- [6] Garrett R G. The chi-square plot: A tool for multivariate outlier recognition [J]. Journal of Geochemical Exploration. 1989, 32: 319-341.
- [7] 杜文, 易建华, 谭新良, 等. 基于近红外光谱的烟叶 SIM-CA 模式识别[J]. 中国烟草学报. 2009, 15(5):1-5.