

氢谱核磁共振与随机森林算法联用快速 检测大豆油过氧化值

荣 菡¹, 杨 丹^{2,3}, 甘露菁¹, 谢 楚⁴

(1. 北京理工大学珠海学院 材料与环 境学院, 广东 珠海 519088; 2. 中 粮 营 养 健 康 研 究 院 营 养 健 康 与 食 品 安 全 北 京 市 重 点 实 验 室, 北 京 102209; 3. 国 贸 食 品 科 技 (北 京) 有 限 公 司, 北 京 102209; 4. 北 京 理 工 大 学 珠 海 学 院 数 理 与 土 木 工 程 学 院, 广 东 珠 海 519088)

摘要: 基于氢谱核磁共振技术与随机森林算法联用, 建立了大豆油过氧化值的快速检测方法。通过对比大豆油氢谱核磁共振数据, 分析鉴别出 5 种氧化产物, 分别为 (*Z, E*) 型和 (*E, E*) 型共轭烯烃类氢过氧化物、直链烷醛、(*E, E*) - 共轭二烯醛、(*E*) - 共轭烯醛。根据氢谱核磁共振测定了油样氧化产物的含量, 同时采用比色法测定了过氧化值, 采用交叉验证的方式构建了大豆油过氧化值快速检测的定量预测模型。结果表明: 建模参数决策树的数量 (ntree) 为 700、决策树的最大深度 (max_depth) 为 5、决策树最小样本节点数量 (min_samples_split) 为 4 时, 所建的预测模型标准均方根偏差 (S_{RME}) 为 2.851%, 模型过氧化值预测值与比色法 (标准方法) 测定结果相关系数 (R^2) 为 0.943, 模型表现出良好的稳定性和预测精度。氢谱核磁共振与随机森林算法联用可快速在线检测大豆油过氧化值, 为油脂品质指标监测提供技术支持。

关键词: 氢谱核磁共振; 过氧化值; 随机森林算法; 定量预测; 大豆油

中图分类号: TS227; O657

文献标识码: A

文章编号: 1003-7969(2021)06-0125-05

Rapid determination of peroxide value of soybean oil using ¹H NMR spectroscopy combined with random forest algorithm

RONG Han¹, YANG Dan^{2,3}, GAN Lujing¹, XIE Chu⁴

(1. School of Materials and Environment, Beijing Institute of Technology, Zhuhai, Zhuhai 519088, Guangdong, China; 2. Beijing Key Laboratory of Nutritional Health and Food Safety, COFCO Nutrition and Health Research Institute, Beijing 102209, China; 3. National Trade Food Technology (Beijing) Co., Ltd., Beijing 102209, China; 4. School of Applied Science and Civil Engineering, Beijing Institute of Technology, Zhuhai, Zhuhai 519088, Guangdong, China)

Abstract: A rapid detection method of peroxide value of soybean oil was established by ¹H nuclear magnetic resonance (¹H NMR) spectroscopy combined with random forest algorithm. By comparing the ¹H NMR data of soybean oil, five kinds of oxidation products were identified, such as (*Z, E*) - conjugated form, (*E, E*) - conjugated form, *E* - alkenals, (*E, E*) - alkadienals and *n* - alkanals. According to the content of five kinds of oxidation product in soybean oil samples determined by the ¹H

收稿日期: 2020-04-12; 修回日期: 2020-09-15

基金项目: 广东高校省级重点平台及重大科研项目 (青年创新人才类) (2017KQNCX250); 北京市科技计划 (Z1711000 01317004)

作者简介: 荣 菡 (1984), 女, 讲师, 硕士, 主要从事食品安全快速检测与功能性食品研究工作 (E-mail) eddie_rong@163.com。

通信作者: 杨 丹, 高级工程师 (E-mail) dyang0628@outlook.com。

NMR data and the peroxide value of soybean oil samples detected by colorimetry, the quantitative prediction model for the rapid detection of peroxide value was constructed by cross-validation. The results showed that the model could present the stronger learning ability and accuracy when the optimization parameters of ntree 700, max_depth 5 and min_samples_split 4, and the model's coefficient (R^2) and root mean

square error (RMSE) were 0.943 and 2.851%, respectively. The method based on ^1H NMR spectroscopy combined with random forest algorithm could determine peroxide value rapidly, which provided scientific support for monitoring and evaluating the oil quality online.

Key words: ^1H nuclear magnetic resonance (^1H NMR); peroxide value; random forest algorithm; quantitative prediction; soybean oil

氢谱核磁共振是基于氢核原子核磁性的一种波谱技术,具有重现性好、操作简便、结构信息丰富等优势,它能够根据氢核在一定的化学环境变化中造成的化学位移,检测并分析与食用油品质相关的脂肪酸、醛类、醇类、环氧化物等指标,在油脂品质检测、产地溯源、掺假检测等方面已有应用^[1-5]。

油脂过氧化值检测通常采用碘量法和比色法,需要大量有机试剂且操作烦琐费时,特别是碘量法,使用的试剂种类、用量、提取方式与时长等因素均会影响检测结果,因此有必要开发更为快速且可靠的食用油过氧化值检测方法。

目前,在我国鲜少有应用核磁共振技术结合机器学习算法快速实现食用油品质指标在线检测的研究。本研究通过氢谱核磁共振,分析了大豆油样品在自动氧化过程中形成的初级、二级氧化产物,并根据氢谱数据分析氧化产物,同时采用比色法检测了大豆油样品的过氧化值,作为建模的定标参考数据。此外,基于随机森林算法能够处理复杂信息、多维数据结构矩阵各变量之间高度相关的特点^[6-7],将氢谱核磁共振与随机森林算法联用,建立一种能够根据氧化产物信息快速输出过氧化值的方法模型,以满足我国油脂监管部门面对大样本量的食用油品质指标快速检测的需要。

1 材料与方法

1.1 实验材料

大豆油,中粮营养健康研究院提供。氘代三氯甲烷(CDCl_3 , 99.8%氘代度,含0.03%四甲基硅烷(TMS)),分析纯,德国默克公司; $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$ 、 $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 、 HCl 、 FeCl_2 、 NH_4SCN 、氯仿-甲醇(体积比7:3),均为分析纯,北京化工厂。

AVANCE 600MHz 超导傅里叶变换核磁共振仪、CPPBBO 超低温探头,瑞士布鲁克公司; HITACHI U-3900 紫外分光光度计,日本日立公司; 5 mm 核磁共振样品管,美国 Norell 公司; TAR-GIN 涡旋仪,北京踏锦科技有限公司。

1.2 实验方法

1.2.1 不同氧化程度大豆油样品的制备

将 50 g 大豆油样品分装入 100 mL 玻璃试管中,放置在设有白炽灯的 40℃ 烘箱中,按照存放时间的

不同,定期取等量样品置于洁净样品瓶中冷冻保存,待分析。共采集不同氧化程度的大豆油样品 340 例。

1.2.2 核磁共振氢谱数据采集

称取待测油样 60 mg 于 2 mL 液相小瓶中,加入 0.6 mL 质量浓度为 2 mg/mL 的氘代三氯甲烷,涡旋 1 min,使样品充分混合,取 0.5 mL 至核磁管中,上机检测。以四甲基硅烷峰为化学位移零点,记录氢谱信号峰值。将 ^1H NMR 信号导入 IconNMR spect 软件,自动调节相位和基线,以氘代三氯甲烷内标峰定标校正化学位移,归一化进行积分。

分析条件:90°脉冲;虚拟扫描次数为 4;扫描次数为 64;谱线宽度 13.993 1;采集时间 5.852 365 s;频率偏移 6.500;温度 25℃。

试样中氧化产物的含量以内标氘代三氯甲烷计算得到,其计算公式为:

$$n_{\text{IS}} = m_{\text{IS}}/M_{\text{IS}} = \rho \times V \times 0.02\% / M_{\text{IS}} \quad (1)$$

$$n_{\text{S}} = A_{\text{S}} \times n_{\text{IS}} / A_{\text{IS}} \quad (2)$$

$$b_{\text{S}} = n_{\text{S}} \times 10^6 / m_{\text{S}} \quad (3)$$

式中: n_{IS} 为内标物质的量, mol; m_{IS} 为样品溶液中内标物质量, g; M_{IS} 为内标物摩尔质量, g/mol; ρ 为样品溶液 25℃ 时的密度, g/L; V 为样品溶液体积, L; n_{S} 为样品中氧化产物物质的量, mol; A_{S} 为目标组分的峰面积; A_{IS} 为内标物峰面积; m_{S} 为样品质量, g; b_{S} 为样品中氧化产物的含量, mmol/kg。

1.2.3 比色法测定过氧化值

取 0.4 g $\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$ 溶于 50 mL 水中,取 0.5 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 溶于 50 mL 水中,将两者缓慢混合并加入 2 mL HCl ,形成 FeCl_2 溶液。取 30 g NH_4SCN 溶于水中,定容至 100 mL,得到 NH_4SCN 溶液。取氧化大豆油样品 0.2 g 于 9.8 mL 氯仿-甲醇(体积比 7:3)中,涡旋 2~4 s 混匀,加入 50 μL FeCl_2 溶液和 50 μL NH_4SCN 溶液,涡旋 2~4 s 混匀后,在室温下放置 5 min,用紫外分光光度计在 500 nm 下测定吸光度。同时,以不加油样的为空白测定吸光度,基于油脂氧化产生的过氧化物与 Fe^{2+} 反应,以 Fe^{3+} 制作标准曲线^[8],按式(4)计算油样的过氧化值,并将其作为模型的标准方法的参考定标数据。

$$X = (A_{\text{s}} - A_{\text{b}}) \times (41.25/55.84) / (m \times 2) \quad (4)$$

式中: X 为样品的过氧化值, mmol/kg; A_s 为样品的吸光度; A_b 为空白样的吸光度; 41.25为标准曲线斜率; 2为换算因子; 55.84为Fe的原子量, m 为样品质量, mg。

1.2.4 样品过氧化值实时检测模型的建立

样品过氧化值实时检测模型的实施, 主要包括两部分: 模型训练和再预测。通过一定数量的大豆油样品, 经由氢谱核磁共振分析其氧化产物, 提取特征作为模型的输入向量, 通过算法, 优化参数并训练模型, 直至模型预测精度良好且稳定。模型建好后, 大批样本的大豆油经氢谱核磁共振扫描后, 相关数据直接代入已建好的预测模型中, 通过机器学习, 即可显示输出过氧化值的预测值, 可为大豆油的质量评估提供参考依据。

建立模型执行内部交叉互验法, 一般用模型的相关系数(R^2)和标准均方根偏差(S_{RME})来评价模型的预测效果和精度。

2 结果与讨论

2.1 确定特征结构的氧化产物

油脂在氧化环境中, 经过链引发、链增殖、链终止等阶段, 生成初级氧化产物——氢过氧化物, 氢过氧化物不稳定, 继续分解生成二级氧化产物——醛类、酮类、烯炔、酸类、醇类等化合物。

通过比对大豆油氢谱核磁共振谱图中信号峰的化学位移与峰的类型归属^[9-10], 得到样品初级氧化产物脂类氢过氧化物的氢谱特征峰信号(B、F)和二级氧化产物中醛的氢谱特征峰信号(I、V、VI), 结果见图1。样品氧化产物的特征基团的核磁信号分配见表1。

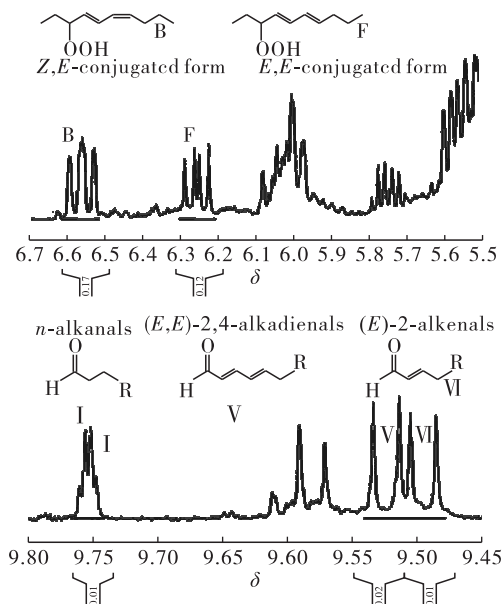


图1 大豆油的氢谱核磁共振谱图

表1 氧化产物的氢谱核磁共振信号分配

峰信号	化学位移	特征基团	
		氢核质子类型	基团特征
B	6.55	—CH=CH—CH=CH—	(Z,E)-共轭双键结构
F	6.25	—CH=CH—CH=CH—	(E,E)-共轭双键结构
I	9.75	—CHO	n-烷醛
V	9.52	—CHO	(E,E)-2,4-烯醛
VI	9.48	—CHO	(E)-2-烯醛

由表1可见, B、F信号分别与不饱和脂肪酸在初级氧化过程中产生的(Z,E)型和(E,E)型共轭烯炔类氢过氧化物中碳碳双键上的氢相关, I、V、VI信号分别与二级氧化过程中产生的直链烷醛、(E,E)-共轭二烯醛和(E)-共轭烯醛中的醛基氢相关。基于此, 可计算得出(Z,E)型和(E,E)型共轭烯炔类、直链烷醛、(E,E)-共轭二烯醛和(E)-共轭烯醛等氧化产物的含量。

将5种氧化产物含量的数据输入Python软件的统计方法源, 经邓肯检验后, 发现具有显著性水平($p < 0.05$), 说明氧化产物的含量与过氧化值密切相关, 因此以氢谱特征峰信号(B、F、I、V、VI)所对应基团特征的5种氧化产物的含量作为输入向量, 建立快速预测过氧化值的定量模型是可行的。

表2是通过Python软件得到的不同氧化产物氢谱特征信号峰与过氧化值相关性贡献率得分。由表2可见, 在不同氧化产物中, 与过氧化值最相关的因子是来自初级氧化产物的结构信号B和F所反映的基团特点, 这是因为当初级氧化产物生成时, 自由基引发的活化能较高, 氧化反应活性较高, 与油脂氧化的进程紧密相关^[11]。

表2 氧化产物与过氧化值相关性贡献率信度得分

氧化产物特征	与过氧化值相关性的贡献率得分
(Z,E)-共轭双键结构	0.897
(E,E)-共轭双键结构	0.871
n-烷醛	0.522
(E,E)-2,4-烯醛	0.557
(E)-2-烯醛	0.514

2.2 随机森林算法构建定量预测模型

随机森林通过重抽样方法从原始样本中抽取多个样本构成样本子集, 利用每个样本子集构建决策树, 执行多棵决策树的预测结果, 正确的剪枝策略是优化算法的核心。建模中需要优化的参数主要有决

策树的数量(ntree)、决策树的最大深度(max_depth)、决策树最小样本节点数量(min_samples_split)。

根据实验需要,最小样本节点数量取值建议从小递增,图2是决策树所需要的最小样本节点数量,分别取值2、4、8优化建模时,并综合决策树的最大深度值,得到的预测模型的评分。

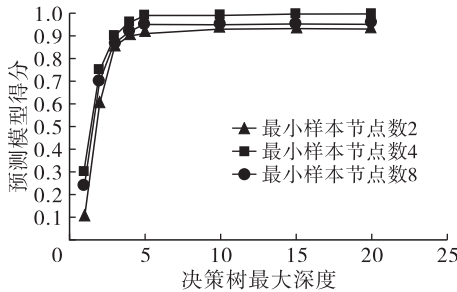


图2 决策树不同最小样本节点数量与模型评分情况

由图2可见,当决策树的最大深度大于5时,模型得分几乎保持恒定,根据决策树的生长原则,避免因决策树扩增导致的样本信息训练不够的情况发生,选择决策树的最大深度值5为建模最优参数之一。同时,尽管从最小样本节点数量取值来看,不同取值调整对模型评分未见明显影响,但决策树最小样本节点数量值为4时,模型评分最高。即根据样本量训练的具体情况以及科学剪枝的原则,使模型分型分支后的子节点保证至少4个训练样本,可让模型更加平滑,交叉验证时拟合效果最佳。

确定合适的决策树的数量对随机森林模型的性能、可解释性和复杂性之间的平衡均有重要意义。决策树的数量取值通常要根据建模数据集的特点而定。决策树数量较小,随机森林的分类和回归误差大,性能较差;决策树数量过大,模型过于复杂且费时,甚至决策树数量大到一定程度时,模型的可解释性减弱。图3是决策树数量对模型的影响。由图3可见,当决策树数量为700时,模型的 S_{RME} 值最小,为2.851%。综上,建模参数如下:决策树的数量为700,决策树的最大深度为5,决策树最小样本节点数量为4。

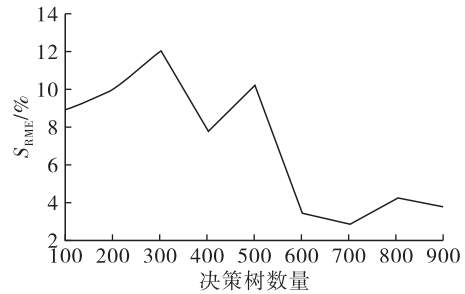


图3 决策树数量对模型的影响

在最佳建模参数下建模,训练集油样过氧化值预测值与比色法测定的油样过氧化值真值的相关性见图4。由图4可见,训练集过氧化值预测值与比色法测定结果一致,相关系数为0.943,说明所建模型预测精度高。

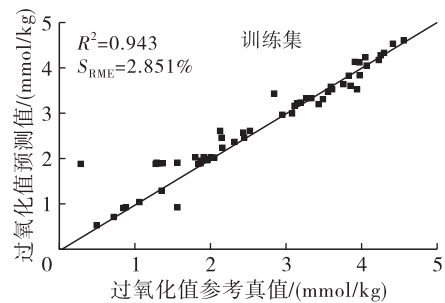


图4 模型预测值与比色法测定值的相关性

2.3 模型预测能力

模型预测能力重现性是衡量所建模型可靠性的重要指标。为检验模型的重现性,对同一未知样品经氢谱核磁共振仪器扫描8次后,随机森林模型预测得到的过氧化值分别为1.3149、1.3111、1.3099、1.3112、1.3098、1.3054、1.3125、1.3056 mmol/kg,标准差为0.0032 mmol/kg,同时采用化学比色法实测值为1.3089 mmol/kg。说明模型的稳定性和重现性令人满意。

表3是基于随机森林算法快速预测过氧化值的预测结果。测试时,为减少在建模时的测量噪声,数据经平滑化处理。由表3可见,过氧化值的定量分析模型中的相对误差较小,说明模型预测结果良好。

表3 随机森林回归模型的预测结果

编号	比色法标准值/ (mmol/kg)	本法预测值/ (mmol/kg)	相对误差/%	编号	比色法标准值/ (mmol/kg)	本法预测值/ (mmol/kg)	相对误差/%
1	2.990 0	3.001 9	0.40	9	0.750 1	0.730 0	-2.68
2	0.600 1	0.640 2	6.68	10	0.600 4	0.610 5	1.68
3	0.660 3	0.620 6	-6.01	11	0.820 1	0.780 2	-4.87
4	0.631 0	0.660 8	4.72	12	0.700 3	0.660 1	-5.74
5	0.710 2	0.690 4	-2.79	13	1.121 0	1.141 0	1.78
6	0.690 7	0.710 6	2.88	14	0.580 1	0.630 2	8.64
7	0.700 1	0.740 0	5.70	15	0.590 3	0.570 0	-3.44
8	0.651 0	0.710 9	9.20	16	1.220 1	1.190 1	-2.46

续表 3

编号	比色法标准值/ (mmol/kg)	本法预测值/ (mmol/kg)	相对误差/ %	编号	比色法标准值/ (mmol/kg)	本法预测值/ (mmol/kg)	相对误差/ %
17	0.940 3	0.890 1	-5.34	37	3.601 1	3.517 5	-2.32
18	0.630 1	0.640 1	1.59	38	3.656 5	3.577 6	-2.16
19	0.681 0	0.641 0	-5.87	39	3.640 1	3.630 3	-0.27
20	4.370 1	4.120 0	-5.72	40	4.101 1	3.773 3	-7.99
21	4.780 2	4.700 1	-1.68	41	3.511 4	3.359 2	-4.33
22	1.881 9	1.896 7	0.79	42	3.900 9	3.648 6	-6.47
23	1.988 9	1.921 9	-3.37	43	3.989 2	3.571 9	-10.46
24	1.909 7	1.972 7	3.30	44	4.023 0	3.887 1	-3.38
25	1.988 0	2.007 1	0.96	45	4.278 3	4.230 6	-1.11
26	2.045 6	2.038 1	-0.37	46	3.900 1	3.653 4	-6.33
27	2.083 3	2.041 7	-2.00	47	4.612 3	4.673 8	1.33
28	2.189 2	2.262 7	3.36	48	4.020 3	4.181 8	4.02
29	2.351 1	2.392 8	1.77	49	4.118 2	4.107 8	-0.25
30	2.467 8	2.599 0	5.32	50	4.101 2	4.293 2	4.68
31	3.401 4	3.373 2	-0.83	51	3.806 6	3.692 1	-3.01
32	3.119 3	3.035 0	-2.70	52	4.305 7	4.339 9	0.79
33	3.157 2	3.204 6	1.50	53	3.477 8	3.241 3	-6.80
34	3.238 6	3.277 3	1.19	54	3.532 5	3.354 1	-5.05
35	3.377 0	3.374 0	-0.09	55	3.299 4	3.255 7	-1.32
36	3.309 3	3.371 9	1.89				

3 结 论

本研究基于氢谱核磁共振技术,通过分析大豆油的氧化产物,采用随机森林算法建立了快速检测大豆油过氧化值的定量预测模型。结果表明,模型稳健且预测结果良好,这与决策树算法的自组织、自学习能力有关。该方法能够为大样本量的食用油品质控制与评价提供一种新思路,具有一定的应用价值。

参考文献:

- [1] ELISABETTA S, GABRIELLA P, GIULIO C, et al. Identification of production chain of Asiagod' Allevo cheese by nuclear magnetic resonance spectroscopy and principal component analysis [J]. J Agric Food Chem, 2008, 56 (16): 7208 - 7214.
- [2] 李玮, 姜洁, 路勇, 等. NMR 氢谱定量测定奶酪中总共轭亚油酸的含量 [J]. 食品科学, 2015, 36(10): 134 - 138.
- [3] ALONSO - SALCES R M, MORENO - ROJAS J M, HOLLAND M V, et al. Virgin olive oil authentication by multivariate analyses of ^1H NMR fingerprints and $\delta^{13}\text{C}$ and $\delta^2\text{H}$ data [J]. J Agric Food Chem, 2010, 58(9): 5586 - 5596.
- [4] D'IMPERIO M, GOBBINO M, PICANZA A, et al. Influence of harvest method and period on olive oil composition; an NMR and statistical study [J]. J Agric Food Chem, 2010, 58(20): 11043 - 11051.
- [5] 石婷, 陈倩, 闫小丽, 等. 核磁共振氢谱结合化学计量学快速检测掺假茶油 [J]. 食品科学, 2018, 39 (22): 241 - 248.
- [6] CHEN L, CHU C, HUANG T, et al. Prediction and analysis of cell - penetrating peptides using pseudo - amino acid composition and random forest models [J]. Amino Acids, 2015, 47(12): 1485 - 1493.
- [7] SINGH B, SIHAG P, SINGH K. Modelling of impact of water quality on infiltration rate of soil by random forest regression [J]. Model Earth Syst Environ, 2017, 56 (3): 999 - 1004.
- [8] SHANTHA N C, DECKER E A. Rapid, sensitive, iron - based spectrophotometric methods for determination of peroxide values of food lipids [J]. J AOAC Int, 1994, 77 (2): 421 - 424.
- [9] NIEVA - ECHEVARRÍA B, GOICOECHEA E, GUILLÉN M D. Behaviour of non - oxidized and oxidized flaxseed oils, as models of ω -3 rich lipids, during in vitro digestion. Occurrence of epoxidation reactions [J]. Food Res Int, 2017, 97: 104 - 115.
- [10] GOICOECHEA E, GUILLEN M D. Analysis of hydroperoxides, aldehydes and epoxides by ^1H nuclear magnetic resonance in sunflower oil oxidized at 70 and 100 $^{\circ}\text{C}$ [J]. J Agric Food Chem, 2010, 58(10): 6234 - 6245.
- [11] GUILLEÉN M D, URIARTE P S. Contribution to further understanding of the evolution of sunflower oil submitted to frying temperature in a domestic fryer; study by ^1H nuclear magnetic resonance [J]. J Agric Food Chem, 2009, 57(17): 7790 - 7799.