

傅里叶变换红外光谱的牛乳中 α_s1 -酪蛋白和 κ -酪蛋白含量的快速检测

肖仕杰¹, 王巧华^{1, 2*}, 樊懿楷³, 刘锐³, 阮健³, 温万⁴,
李季奇⁴, 邵怀峰⁴, 刘维华⁵, 张淑君^{3*}

1. 华中农业大学工学院, 湖北 武汉 430070
2. 农业部长江中下游农业装备重点实验室, 湖北 武汉 430070
3. 华中农业大学动物遗传育种与繁殖教育部实验室, 湖北 武汉 430070
4. 宁夏回族自治区畜牧工作站, 宁夏 银川 750002
5. 宁夏回族自治区兽药饲料监察所, 宁夏 银川 750011

摘要 为了找到一种能够对牛乳中的两种主要过敏原(α_s1 和 κ -酪蛋白)含量快速检测的方法,以河南、湖北、宁夏和内蒙古四省区的211份中国荷斯坦牛牛乳样本为研究对象,建立了基于傅里叶变换中红外光谱技术的牛乳中 α_s1 和 κ -酪蛋白含量的无损快速检测模型。首先对牛乳的原始光谱进行预分析,发现水对牛乳的光谱吸收具有很强的干扰,对水的两个主要吸收区域 $1597\sim 1712$ 和 $3024\sim 3680\text{ cm}^{-1}$ 进行分析,发现水的吸收区域 $1597\sim 1712\text{ cm}^{-1}$ 和蛋白的部分吸收区域 $1558\sim 1705\text{ cm}^{-1}$ (酰胺I)基本重合,通过对比去除 $1597\sim 1712\text{ cm}^{-1}$ 前后的效果,最终选择 $925.92\sim 3005.382\text{ cm}^{-1}$ 的光谱区域作为敏感波段用于后续分析。选取的全光谱经手动降维,利用MCCV剔除异常样本,分别采用标准正态变量变换(SNV)、多元散射校正(MSC)等8种预处理算法和竞争性自适应重加权算法(CARS)、无信息变量消除法(UVE)等3种特征选择算法联合建立支持向量机回归模型(SVR)。经检验,对于 α_s1 -酪蛋白,一阶导数和CARS算法结合建立的SVR模型效果最优,训练集相关系数 R_c 和测试集相关系数 R_p 分别为0.8827和0.8998,训练集均方根误差RMSEC和测试集均方根误差RMSEP分别为1.1363和1.3726;对于 κ -酪蛋白,一阶差分法和UVE算法结合建立的SVR模型效果最优,训练集相关系数 R_c 和测试集相关系数 R_p 分别为0.8808和0.8903,训练集均方根误差RMSEC和测试集均方根误差RMSEP分别为0.5345和0.5354。研究结果表明,基于傅里叶变换中红外光谱技术建立的SVR模型可以对牛乳中的过敏原 α_s1 和 κ -酪蛋白含量进行无损检测,预测效果良好,此研究弥补了国内利用光谱技术对牛乳中的酪蛋白进行无损快速检测的空白。

关键词 中红外光谱;牛乳; α_s1 -酪蛋白; κ -酪蛋白;无损检测

中图分类号: S24 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)12-3688-07

引言

牛乳中含有丰富的蛋白质,对人体的生长发育起着重要作用,尤其对于婴幼儿来说是不可或缺的优质蛋白质来源,但同时,牛乳也是一种过敏原。FAO/WHO已经将牛奶和乳制品确定为引发人类食物过敏现象的8种主要食物之一^[1],相关数据表明牛乳过敏患病率在婴儿中高达2%~7.5%^[2],

随着乳制品销量的增加,牛乳过敏率不断上涨已变为不可忽略的食品安全问题。对牛乳过敏,实际上是对牛乳中的蛋白质敏感,乳蛋白的两个主要类别分别是乳清蛋白和酪蛋白,其中酪蛋白的含量约占总蛋白质含量的80%,约65%的牛乳过敏人员对酪蛋白过敏。其中, α_s1 和 κ -酪蛋白为主要的过敏原^[3]。牛乳过敏目前无法根治,只能避免饮用牛奶或食用乳制品。牛乳蛋白过敏病人通常在消化系统和皮肤两个方面有明显的症状表现,如呕吐、腹泻、腹痛和湿疹、荨麻疹

收稿日期:2020-11-23,修订日期:2021-02-18

基金项目:欧盟FP7构架项目(FP7-KBBE-2013-7-613689)资助

作者简介:肖仕杰,1993年生,华中农业大学工学院硕士研究生 e-mail:1175760869@qq.com

*通讯作者 e-mail:wqh@mail.hzau.edu.cn;sjxiaozhang@mail.hzau.edu.cn

等^[4], 因此很多国家都制定了食品过敏原强制标识条例来保障大众健康^[3], 我国制定的 GB 7718—2011《预包装食品标签通则》^[5]和 GB/T 23779—2009《预包装食品中的致敏原成分》^[6]建议商家标明可能的致敏物。

如果能够可靠地检测出牛乳中 α_s1 和 κ -酪蛋白的含量, 就能为牛乳敏感人员提供饮用参考指示。乳成分的主要检测方法有气相色谱法、色谱-质谱联用法和高效液相色谱法等, 这些方法灵敏度高、可靠性好, 但成本高、技术难度大、分析时间长, 因此找到一种简单高效的替代方法非常重要。红外光谱法具有快速无损、简单易行的优点, 相比于近红外光谱, 中红外光谱的波段范围更广、包含的信息量更丰富, 在国外被广泛应用于牛乳中各营养成分如蛋白成分的检测。Etzion 等^[7]表明中红外光谱法可以预测乳蛋白的含量, Bonfatti 等^[8]基于中红外光谱对牛乳中的酪蛋白等的含量进行了预测, Niero 等^[9]表明 UVE 算法可以提高中红外光谱对乳蛋白组分含量的预测精度, McDermott 等^[10]基于中红外光谱预测了牛乳中的蛋白质和氨基酸含量。但在国内, 利用光谱技术检测牛乳中蛋白成分的研究鲜有报道。

为此, 本文利用傅里叶变换中红外光谱技术对牛乳中 α_s1 和 κ -酪蛋白两种过敏原进行分析, 利用竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS)、无信息变量消除法 (uninformative variables elimination, UVE) 和连续投影算法 (successive projections algorithm, SPA) 筛选出能代表酪蛋白含量的特征变量, 并利用支持向量机 (support vector regression, SVR) 模型分别构建了 α_s1 -酪蛋白含量和 κ -酪蛋白含量的无损检测模型, 模型的预测精度优于 Bonfatti 等^[8]、Niero 等^[9]和 McDermott 等^[10]前人研究结果。

1 实验部分

1.1 试验材料

试验材料来源于河南、湖北、宁夏和内蒙古四省区的 211 头中国荷斯坦牛, 一头牛采集一份牛乳, 牛乳采集利用自动挤奶装置完成, 先用消毒过的毛巾擦拭牛乳房, 然后用碘甘油混合溶液再次消毒, 挤出前三把乳汁后进行牛乳采集, 每份牛乳采集 40 mL, 分装到直径 3.5 cm, 高 9 cm 的圆柱形全新采样瓶里, 依次编号, 并向每个采样瓶里立即加入溴丙二醇防腐剂, 缓慢摇晃使其充分溶解, 运回途中在牛乳样品周围放置冰袋防止变质, 到达实验室后立即放入冰箱保存 (4 °C), 并于第二天进行光谱采集。

1.2 仪器、设备和试剂

MilkoScan™ FT+ [傅里叶变换中红外光谱仪 (FTIR), 丹麦 FOSS 公司]; 电热恒温水浴锅; 十万分之一电子天平; Waters e2695 液相色谱仪。

α_s1 -酪蛋白 (lot C-6780, 纯度 $\geq 70\%$)、 κ -酪蛋白 (lot C-0406, 纯度 $\geq 80\%$) 标准品 (Sigma 公司); 乙腈 (色谱级, 纯度 $\geq 99.8\%$)、盐酸胍、三氟乙酸 (TFA) (上海生工公司); 其他试剂均为国产分析纯。

1.3 方法

1.3.1 中红外光谱的采集

利用 MilkoScan™ FT+ 进行光谱采集, 具体采集步骤: 将牛乳分批放入 45 °C 电热恒温水浴锅内预热 5 min, 预热的牛乳放在检测架上上下摇晃数次使牛乳混合均匀, 将检测架放在检测履带上, 打开瓶盖, 依次进行检测, 采集完光谱后的牛乳置于 -20 °C 冷冻保存。

1.3.2 α_s1 -酪蛋白和 κ -酪蛋白的含量测定

(1) 标准样品的处理

先用去离子水将混合标样溶解, 直到浓度约为 $10 \text{ g} \cdot \text{L}^{-1}$ 左右, 然后往 1 600 μL 处理液 ($6 \text{ mol} \cdot \text{L}^{-1}$ 盐酸胍溶液, 内含 $0.1 \text{ mmol} \cdot \text{L}^{-1}$ EDTA-Na₂, pH 6.0) 中滴加 400 μL 配好的标样溶液, 于室温下孵育 90 min, 上机前用 0.22 μm 尼龙滤膜过滤。

(2) 牛乳的处理

取 80 μL 牛乳于 320 μL 处理液中, 室温孵育 90 min, 将离心机转速调为 $14\,000 \text{ r} \cdot \text{min}^{-1}$, 5 min 后取上清液。上机前用 0.22 μm 尼龙滤膜过滤。

(3) RP-HPLC 的色谱条件

色谱柱: ZORBAX 300SB-C18; 进样量: 50 μL ; 柱温: 40 °C; 流速: $1 \text{ mL} \cdot \text{min}^{-1}$; 洗脱时间: 42 min; 检测波长: 214 nm。

流动相 A: 10% 乙腈 + 90% 去离子水 + 0.1% TFA; B: 90% 乙腈 + 10% 去离子水 + 0.1% TFA。流动相 B 梯度 (变化率) 如下: 从 33% 到 38% 洗脱 10 min ($0.50\% \text{ B} \cdot \text{min}^{-1}$), 从 38% 到 40% 洗脱 6 min (约 $0.33\% \text{ B} \cdot \text{min}^{-1}$), 保持 40% 洗脱 6 min ($0.00\% \text{ B} \cdot \text{min}^{-1}$), 从 40% 到 40.5% 洗脱 2 min ($0.25\% \text{ B} \cdot \text{min}^{-1}$), 保持 40.5% 洗脱 2 min ($0.00\% \text{ B} \cdot \text{min}^{-1}$), 从 40.5% 到 48% 洗脱 14 min (约 $0.54\% \text{ B} \cdot \text{min}^{-1}$), 最后立刻以初始梯度平衡色谱柱 2 min, 准备下一批牛乳的检测, 平均每批次检测牛乳 30 份。

同一批次检测结束后用 10% 甲醇 + 90% 去离子水与 100% 甲醇清洗色谱柱, 以保证下一批次牛乳的正常检测。

1.4 光谱预处理、特征提取

牛乳胶束的散射以及仪器运行过程中产生的随机噪声会对光谱造成干扰, 因此光谱中不仅包含许多有用的化学信息, 还存在大量的背景噪声和无用信息。为了最大可能的削弱干扰信息, 保留有效信息, 提高模型的稳健性, 正式建模前先对光谱预处理。分别利用标准正态变量变换 (standard normal variate transformation, SNV)、多元散射校正 (multivariate scatter correction, MSC)、一阶导数、一阶差分、归一化 (normalize)、二阶导数和二阶差分 7 种方法进行预处理。

中红外光谱的波段范围广, 冗余信息繁多, 通过特征提取算法, 能够大大减少光谱维数, 优化算法, 提高模型的识别率。本文利用 CARS、UVE 和 SPA 算法提取特征变量。

CARS 算法^[11]基于“优胜劣汰”准则剔除不适应的波长变量, 在有效去除无信息变量的同时压缩共线性变量, 最终选择出针对预测目标最为关键的变量。

UVE 算法^[12]基于 PLS 回归系数进行变量选择, 该算法的基本思想是利用回归系数来衡量变量的权重, 消除模型中低贡献率的特征变量。

SPA 算法^[13]是一种让变量间共线性最小化的算法, 能

cross-validation, MCCV)对 α_s 1 和 κ -酪蛋白分别进行奇异样本检测与剔除。MCCV 基于 PLS 获取最佳主成分数, 利用随机数按 4 : 1 的原则将光谱数据和酪蛋白测定值划分为训练集和测试集, 分别建立 PLS 回归模型, 设定循环次数为 2 500, 计算出各牛乳的预测残差后分别求均值与方差^[16], 结果如图 3 所示, α_s 1-酪蛋白模型的奇异样本编号为 39 号、75 号和 141 号, κ -酪蛋白模型的奇异样本编号为 61 号、75 号、76 号、141 号和 144 号。

2.3 样本划分

SPXY(sample set partitioning based on joint X-Y distances)法在划分样本时同时考虑了光谱数据和测定的理化指标, 被划分的样本集更合理^[11], 本文利用 SPXY 将剔除异常后的样本按 7 : 3 划分为训练集和测试集。其中, α_s 1-酪蛋白的训练集和测试集样本数量分别为 146 和 62, κ -酪蛋白的训练集和测试集样本数量分别为 145 和 61, 各样本集的数据统计情况如表 1 所示。

表 1 利用 SPXY 算法划分样本集的数据统计

Table 1 Data statistics of partitioning sample sets by SPXY algorithm

| 指标 | α_s 1-酪蛋白/(g · L ⁻¹) | | κ -酪蛋白/(g · L ⁻¹) | |
|-----|---|----------|--------------------------------------|---------|
| | 训练集 | 测试集 | 训练集 | 测试集 |
| 最小值 | 2.783 4 | 2.695 3 | 1.156 1 | 1.158 4 |
| 最大值 | 13.782 4 | 16.208 7 | 6.475 4 | 5.992 9 |
| 平均值 | 7.672 4 | 8.047 0 | 3.171 2 | 2.881 9 |
| 标准差 | 2.407 7 | 2.928 5 | 1.127 5 | 1.125 0 |

2.4 光谱预处理和特征变量选择

在导数预处理中, 利用 Savitzky-Golay 求导法进行 9 点平滑、3 点差分宽度的导数预处理。使用 CARS, UVE 和 SPA 分别对预处理后的光谱数据进行特征提取, 分别找出能够代表 α_s 1-酪蛋白含量与 κ -酪蛋白含量的特征变量。因为 α_s 1 和 κ -酪蛋白的特征变量选择过程相同, 下文仅以 α_s 1-酪蛋白的特征变量选择为例分别对 CARS、UVE 和 SPA 的变量选择过程进行阐述。

(1)CARS 进行变量选择的过程如图 4 所示, 将 CARS 的采样次数设为 50, 采用 5 折交叉验证, 重采样率为 0.8。图 4(a)表明, 随着取样运行次数的增加, 被选取的特征变量数量在逐步减少。图 4(b)的均方根误差(RMSECV)值先逐渐减小, 表明无用信息被消除, 再逐渐增加, 表明有效信息被消除。图 4(c)竖线处迭代 22 次, 取得最小 RMSECV 值。

(2)UVE 进行变量选择的过程如图 5 所示, 将 UVE 的阈值参数设为 0.99, 结合 20 个主成分建立 PLS 模型进行变量选择。图中左侧曲线表示实变量, 右侧曲线表示添加的随机变量, 两条水平虚线为随机变量的最大阈值线, 两条水平线之间为被剔除的非有用变量, 水平线之外则为建模的特征变量, 共选出 108 个变量组合。

(3)SPA 进行变量选择的过程如图 6 所示, 根据 RMSE 的变化来确定被选取的特征变量。在变量个数的增加过程中, RMSE 先迅速下降, 说明光谱中的无效信息被高效剔

除, 然后趋于平稳, 说明无效信息基本有效剔除, 选择此过渡点处的变量作为被选取的特征变量组合。

2.5 模型建立与分析

将不同的预处理和特征选择算法结合获得的特征变量组合分别带入 SVR 模型, α_s 1-酪蛋白的预测结果如表 2 所示。对于 CARS 算法, 一阶导数预处理的光谱数据带入 SVR 模型取得了最优效果, 选择的特征变量数量为 24, 占建模光谱

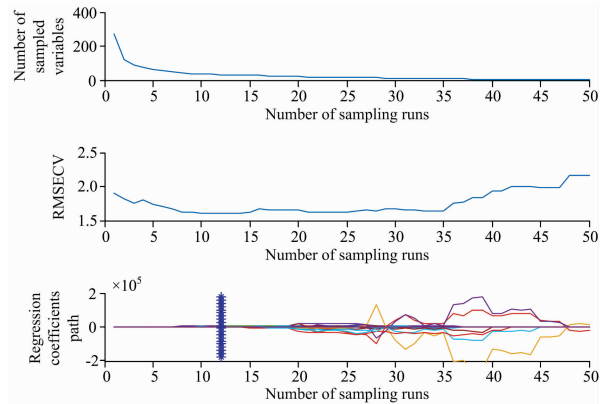


图 4 CARS 变量选择

Fig. 4 Variable selection of CARS

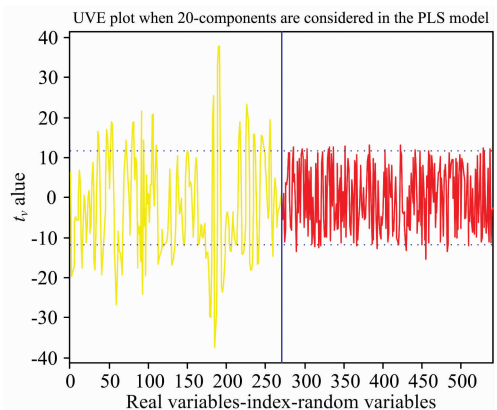
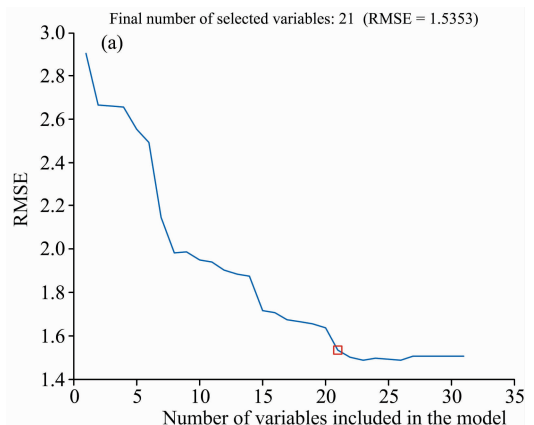


图 5 UVE 消除算法筛选特征波长

Fig. 5 Screening characteristic wavelengths by UVE



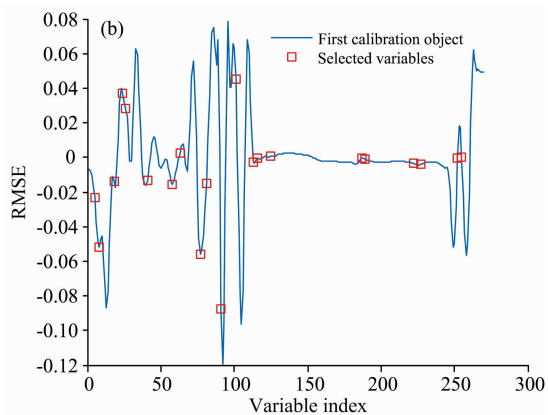


图 6 (a)RMSE; (b)选取的最优波长编号索引

Fig. 6 (a) RMSE; (b) Selected optimal wavelength number index

变量的 8.89%；对于 UVE 算法，一阶导数预处理的光谱数据代入 SVR 模型取得了最优效果，选择的特征变量数量为 108，占建模光谱变量的 40%；对于 SPA 算法，直接将建模光谱带入 SVR 模型取得了最优效果，选择的特征变量数量为 23，占建模光谱变量的 8.52%。

表 2 基于 3 种特征选择算法建立的 α_1 -酪蛋白 SVR 预测模型

Table 2 SVR prediction model of α_1 -casein based on 3 characteristic variable selection methods

| 波长选择算法 | 预处理方式 | 变量个数 | 训练集 | | 测试集 | |
|--------|----------|------|---------|---------|---------|---------|
| | | | R_c | RMSEC | R_p | RMSEP |
| CARS | 原始 | 15 | 0.814 7 | 1.400 6 | 0.756 6 | 1.933 9 |
| | SNV | 38 | 0.837 4 | 1.343 7 | 0.820 9 | 1.771 1 |
| | MSC | 24 | 0.825 1 | 1.402 8 | 0.839 7 | 1.825 2 |
| | 一阶导数 | 24 | 0.882 7 | 1.136 3 | 0.899 8 | 1.372 6 |
| | 二阶导数 | 32 | 0.844 5 | 1.324 3 | 0.800 9 | 1.808 5 |
| | normaliz | 18 | 0.801 9 | 1.447 9 | 0.794 3 | 1.825 6 |
| | 一阶差分 | 11 | 0.833 6 | 1.341 0 | 0.817 8 | 1.746 2 |
| | 二阶差分 | 15 | 0.877 7 | 1.150 8 | 0.767 9 | 1.883 3 |
| UVE | 原始 | 69 | 0.825 7 | 1.370 1 | 0.838 8 | 1.644 1 |
| | SNV | 129 | 0.832 7 | 1.365 5 | 0.816 9 | 1.808 8 |
| | MSC | 137 | 0.812 4 | 1.420 0 | 0.800 6 | 1.887 5 |
| | 一阶导数 | 108 | 0.880 3 | 1.165 3 | 0.851 1 | 1.594 7 |
| | 二阶导数 | 100 | 0.852 6 | 1.310 5 | 0.845 2 | 1.701 5 |
| | normaliz | 101 | 0.840 7 | 1.327 9 | 0.819 1 | 1.746 1 |
| | 一阶差分 | 95 | 0.866 4 | 1.246 6 | 0.835 3 | 1.711 8 |
| | 二阶差分 | 166 | 0.941 2 | 0.840 3 | 0.821 4 | 1.711 8 |
| SPA | 原始 | 23 | 0.825 6 | 1.388 2 | 0.839 2 | 1.676 9 |
| | SNV | 44 | 0.923 7 | 0.929 1 | 0.724 4 | 2.003 4 |
| | MSC | 14 | 0.694 3 | 1.741 7 | 0.794 4 | 1.849 2 |
| | 一阶导数 | 21 | 0.827 0 | 1.381 7 | 0.821 5 | 1.767 8 |
| | 二阶导数 | 21 | 0.943 6 | 0.808 4 | 0.837 1 | 1.603 8 |
| | normaliz | 40 | 0.777 6 | 1.532 1 | 0.743 1 | 1.974 3 |
| | 一阶差分 | 9 | 0.888 2 | 1.129 3 | 0.827 0 | 1.694 1 |
| | 二阶差分 | 6 | 0.798 8 | 1.465 3 | 0.831 0 | 1.707 0 |

κ -酪蛋白的预测结果分别如表 3 所示。对于 CARS 算法，二阶差分预处理的光谱数据带入 SVR 模型取得了最优效果，选择的特征变量数量为 16，占建模光谱变量的 5.93%；对于 UVE 算法，一阶差分预处理的光谱数据带入 SVR 模型取得了最优效果，选择的特征变量数量为 55，占建模光谱变量的 20.37%；对于 SPA 算法，一阶导数预处理的光谱数据带入 SVR 模型取得了最优效果，选择的特征变量数量为 14，占建模光谱变量的 5.19%。

表 3 基于不同特征选择算法建立的 κ -酪蛋白 SVR 预测模型
Table 3 SVR prediction model of κ -casein based on different characteristic variable selection methods

| 波长选择算法 | 预处理方式 | 变量个数 | 训练集 | | 测试集 | |
|--------|----------|------|---------|---------|---------|---------|
| | | | R_c | RMSEC | R_p | RMSEP |
| CARS | 原始 | 33 | 0.883 6 | 0.529 5 | 0.792 5 | 0.685 4 |
| | SNV | 15 | 0.824 5 | 0.649 4 | 0.793 6 | 0.692 4 |
| | MSC | 20 | 0.845 5 | 0.605 6 | 0.826 9 | 0.651 3 |
| | 一阶导数 | 15 | 0.872 4 | 0.552 8 | 0.837 2 | 0.614 1 |
| | 二阶导数 | 15 | 0.885 2 | 0.524 4 | 0.761 4 | 0.726 0 |
| | normaliz | 20 | 0.852 2 | 0.595 8 | 0.823 9 | 0.657 0 |
| | 一阶差分 | 11 | 0.858 2 | 0.580 8 | 0.850 7 | 0.606 6 |
| | 二阶差分 | 16 | 0.864 3 | 0.566 1 | 0.873 9 | 0.576 0 |
| UVE | 原始 | 145 | 0.842 7 | 0.610 1 | 0.828 9 | 0.637 0 |
| | SNV | 107 | 0.893 9 | 0.507 8 | 0.848 7 | 0.596 6 |
| | MSC | 114 | 0.881 0 | 0.538 8 | 0.848 6 | 0.605 4 |
| | 一阶导数 | 57 | 0.860 3 | 0.591 0 | 0.851 4 | 0.618 9 |
| | 二阶导数 | 67 | 0.881 4 | 0.541 5 | 0.857 9 | 0.591 3 |
| | normaliz | 120 | 0.860 8 | 0.576 2 | 0.843 6 | 0.612 8 |
| | 一阶差分 | 55 | 0.880 8 | 0.534 5 | 0.890 3 | 0.535 4 |
| | 二阶差分 | 107 | 0.914 7 | 0.473 5 | 0.887 7 | 0.558 1 |
| SPA | 原始 | 23 | 0.850 4 | 0.593 4 | 0.848 9 | 0.608 9 |
| | SNV | 17 | 0.886 2 | 0.526 3 | 0.834 1 | 0.622 4 |
| | MSC | 15 | 0.861 7 | 0.576 8 | 0.820 3 | 0.638 3 |
| | 一阶导数 | 14 | 0.853 7 | 0.587 7 | 0.885 1 | 0.540 8 |
| | 二阶导数 | 7 | 0.807 6 | 0.664 0 | 0.850 8 | 0.618 1 |
| | normaliz | 9 | 0.825 9 | 0.639 1 | 0.859 1 | 0.608 1 |
| | 一阶差分 | 9 | 0.892 1 | 0.513 0 | 0.808 2 | 0.668 1 |
| | 二阶差分 | 7 | 0.807 6 | 0.664 0 | 0.850 8 | 0.618 1 |

2.6 最优模型的比较

对于 α_1 -酪蛋白预测模型，CARS 算法与 UVE 算法建立的 SVR 模型训练集 R_c 和测试集 R_p 均在 0.85 以上，SPA 算法建立的 SVR 模型训练集 R_c 和测试集 R_p 在 0.82 和 0.85 之间，一阶导数预处理和 CARS 算法结合建立的 SVR 模型最优，训练集 R_c 和测试集 R_p 分别为 0.882 7 和 0.899 8，训练集 RMSEC 和测试集 RMSEP 分别为 1.136 3 和 1.372 6。对于 κ -酪蛋白，CARS 算法、UVE 算法和 SPA 算法建立的 SVR 模型训练集 R_c 和测试集 R_p 均在 0.85 以上，一阶差分预处理和 UVE 算法结合建立的 SVR 模型最优，训练集 R_c 和测试集 R_p 分别为 0.880 8 和 0.890 3，训练集 RMSEC 和测试集 RMSEP 分别为 0.534 5 和 0.535 4。分别将 α_1 和 κ -酪蛋白含量的最优 SVR 回归模型用散点图表

示, 预测结果如图 7 和图 8 所示。

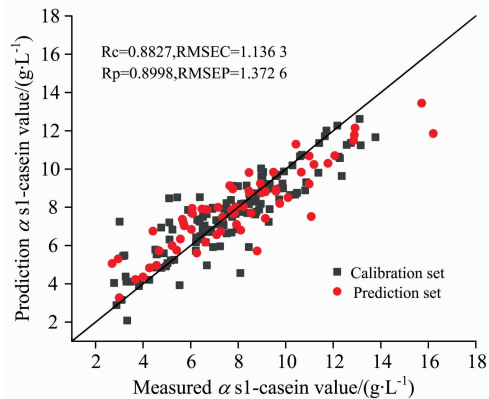


图 7 基于 CARS 的 α_1 -酪蛋白最优模型

Fig. 7 Optimal model for α_1 -casein based on CARS

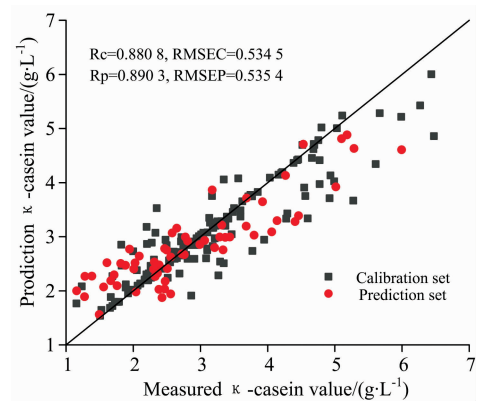


图 8 基于 UVE 的 κ -酪蛋白最优模型

Fig. 8 Optimal model for κ -casein based on UVE

3 结 论

基于傅里叶变换中红外光谱技术, 分别建立了牛乳中 α_1 和 κ -酪蛋白含量的 SVR 无损快速检测模型。

对于 α_1 -酪蛋白, 一阶导数结合 CARS 算法、一阶导数结合 UVE 算法和原始光谱结合 SPA 算法的最优模型提取的特征变量数分别为 24, 108 和 23。结果表明, α_1 -酪蛋白含量的最佳预测模型为一阶导数与 CARS 算法结合建立的 SVR 回归模型, 训练集 R_c 和 RMSEC 分别为 0.8827 和 1.1363, 测试集 R_p 和 RMSEP 分别为 0.8998 和 1.3726; UVE 算法提取的特征变量包含无效信息, 影响了预测精度;

CARS 算法与 SPA 算法提取的特征变量数相当, 但 SPA 算法的精度远低于 CARS 算法, 表明 SPA 算法不适合 α_1 -酪蛋白含量预测模型的建立。

对于 κ -酪蛋白, 二阶差分结合 CARS 算法、一阶差分结合 UVE 算法和一阶导数结合 SPA 算法的最优模型提取的特征变量数分别为 16, 55 和 14; 结果表明, κ -酪蛋白含量的最佳预测模型为一阶差分与 UVE 算法结合建立的 SVR 模型, 训练集 R_c 和 RMSEC 分别为 0.8808 和 0.5345, 测试集 R_p 和 RMSEP 分别为 0.8903 和 0.5354。三种算法的预测精度较为接近, UVE 算法优于 CARS 算法与 SPA 算法, 表明 κ -酪蛋白含量最佳预测模型的建立需要提取更多的特征变量。

本研究可为后续利用中红外光谱法对牛乳中其他与过敏有关的蛋白含量进行快速无损检测提供重要参考。

References

- [1] ZHANG Ai-lin, DUAN Xiao-yun, REN Fei, et al(张爱琳, 段筱筠, 任 斐, 等). Chinese Dairy Industry(中国乳品工业), 2016, 44(12): 4.
- [2] Sicherer S H. Food Allergy. Journal of Allergy and Clinical Immunology, 2011, 125(2): S116.
- [3] Huang H W, Hsu C P, Yang B B, et al. Comprehensive Reviews in Food Science and Food Safety, 2014, 13(1): 78.
- [4] WANG Jia-rui, LI Zhao-xu, LI Shu-guo(王佳蕊, 李朝旭, 李书国). Chinese Dairy Industry(中国乳品工业), 2019, 47(6): 28.
- [5] Ministry of Health of the People's Republic of China(中华人民共和国卫生部). GB 7718—2011 National Food Safety Standard, Rules for the Labeling of Prepackaged Foods(GB 7718—2011 食品安全国家标准 预包装食品标签通则), 2011.
- [6] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China(中华人民共和国国家质量监督检验检疫总局). GB/T 23779—2009 Allergenic Ingredients in Prepackaged Foods(GB/T 23779—2009 预包装食品中的致敏原成分), 2009.
- [7] Etzion Y, Linker R, Cogan U, et al. J. Dairy Sci., 2004, 87: 2779.
- [8] Bonfatti V, Di Martino G, Carnier P. J. Dairy Sci., 2011, 94: 5776.
- [9] Niero G, Penasa M, Gottardo, et al. J. Dairy Sci., 2016, 99: 1853.
- [10] McDermott A, Visentin G, De Marchi M, et al. J. Dairy Sci., 2016, 99: 3171.
- [11] GAO Sheng, WANG Qiao-hua, FU Dan-dan, et al(高 升, 王巧华, 付丹丹, 等). Acta Optica Sinica(光学学报), 2019, 10(39): 1.
- [12] FU Dan-dan, WANG Qiao-hua, GAO Sheng, et al(付丹丹, 王巧华, 高 升, 等). Chinese Journal of Analytical Chemistry(分析化学), 2020, 48(2): 289.
- [13] ZHANG Ben-hua, QIAN Chang-qian, JIAO Jin-kang, et al(张本华, 钱长钱, 焦晋康, 等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2019, 35(18): 237.
- [14] YANG Ren-jie, LIU Rong, XU Ke-xin(杨仁杰, 刘 蓉, 徐可欣). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2011, 31(9): 2383.

[15] XU Feng, FU Dan-dan, WANG Qiao-hua, et al(许 锋, 付丹丹, 王巧华, 等). Food Science(食品科学), 2018, 39(8): 149.

Rapid Determination of α_s1 -Casein and κ -Casein in Milk Based on Fourier Transform Infrared Spectroscopy

XIAO Shi-jie¹, WANG Qiao-hua^{1, 2*}, FAN Yi-kai³, LIU Rui³, RUAN Jian³, WEN Wan⁴, LI Ji-qi⁴, SHAO Huai-feng⁴, LIU Wei-hua⁵, ZHANG Shu-jun^{3*}

1. College of Engineering, Huazhong Agricultural University, Wuhan 430070, China

2. Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River; Ministry of Agriculture and Rural Agriculture, Wuhan 430070, China

3. Key Laboratory of Animal Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China

4. Ningxia Hui Autonomous Region Animal Husbandary Workstation, Yinchuan 750002, China

5. Ningxia Hui Autonomous Region Veterinary Medicine and Feed Supervision Institute, Yinchuan 750011, China

Abstract In order to find a rapid detection method for the content of two main allergens (α_s1 and κ -casein) in milk, 211 Chinese Holstein milk samples from four provinces of Henan, Hubei, Ningxia and Inner Mongolia were selected as the research objects, and a non-destructive and rapid detection model of α_s1 and κ -casein in milk was established based on Fourier transform mid infrared spectroscopy. Firstly, the original spectrum of milk was pre-analyzed, and it was found that strongly influenced the spectral absorption of milk. The two main absorption regions of water (1 597~1 712 and 3 024~3 680 cm^{-1}) were analyzed. It was found that the absorption region of water (1 597~1 712 cm^{-1}) overlapped with that of protein (1 558~1 705 cm^{-1})(amide I). By comparing the effect of removing 1 597~1 712 cm^{-1} , the spectral region of 925.92~3 005.382 cm^{-1} was selected as the sensitive band for subsequent analysis. The dimension of the selected full spectrum was reduced manually, and MCCV eliminated the abnormal samples. The support vector machine regression model (SVR) was established by using eight preprocessing algorithms, such as Savitzky-Golay convolution smoothing (S-G), standard normal variable (SNV). Meanwhile, three feature selection algorithms were combined, such as competitive adaptive reweighting algorithm (CARS) and information-free variable elimination algorithm (UVE). The results showed that for α_s1 -casein, the SVR model established by the combination of the first derivative and CARS algorithm was the best, the training set correlation coefficient (R_C) and test set correlation coefficient (R_P) were 0.882 7 and 0.899 8, respectively, and the training set root mean square error (RMSEC) and test set root mean square error (RMSEP) were 1.136 3 and 1.372 6, respectively. For κ -casein, the SVR model established by the combination of second-order difference and UVE algorithm was the best. The training set correlation coefficient (R_C) and test set correlation coefficient (R_P) were 0.914 7 and 0.887 7, respectively, and the training set root mean square error (RMSEC) and test set root mean square error (RMSEP) were 0.473 5 and 0.558 1, respectively. The results showed that the SVR model based on Fourier transform mid-infrared spectroscopy can be used to detect the content of allergens α_s1 and κ -casein in milk, and the prediction effect was good. This study can make up for the blank of rapid and non-destructive detection of casein in milk by spectral technology in China.

Keywords Medium infrared spectrum; Cow's milk; α_s1 -casein; κ -casein; Nondestructive testing

(Received Nov. 23, 2020; accepted Feb. 18, 2021)

* Corresponding authors