

# 基于 IABC-SVR 算法的拉曼光谱定量分析山羊血清蛋白含量

付兴虎, 赵 飞, 王振兴, 芦 鑫, 付广伟, 金 娃, 毕卫红

燕山大学信息科学与工程学院, 河北省特种光纤与光纤传感重点实验室, 河北 秦皇岛 066004

**摘 要** 提出了一种基于拉曼光谱和改进人工蜂群算法优化支持向量机回归(IABC-SVR)算法快速定量检测山羊血清蛋白含量的方法。传统人工蜂群算法在数据区域规模较大时, 收敛速度逐渐减慢, 出现效率低、精准度下降、局部最优解概率高等问题。所提出的算法解决了这些问题, 使算法在进化前期避免陷入局部最优解, 在进化中后期能够保持解的全局搜索能力。常规测定血清蛋白总量的方法通常采用凯氏定氮法、双缩脲法等, 但存在时效慢、污染样本等缺点。采用拉曼光谱法进行检测, 具有快速、无损的优点。以山羊血清为分析对象, 按一定体积比配置 35 组待测样本, 用拉曼光谱仪采集拉曼光谱, 光谱采集范围为  $300\sim 1\ 300\text{ cm}^{-1}$ , 采用基线矫正去除荧光背景, 使用 Savitzky-Golay 光谱平滑法对原始光谱进行平滑处理, 归一化处理光谱数据, 并对拉曼光谱特征峰进行归属。实验结果表明, 拉曼光谱能够表征血清中主要化学集团的信息, 且由于官能团浓度差异, 光谱特征峰强度随浓度变化明显, 因此基于特征峰信息可以测定血清蛋白总量。实验中, 以购买的山羊血清蛋白含量为基准, 通过配置样本的体积比得到各组待测血清样本的蛋白含量, 配置的单个液体样本体积为 3 mL, 随机选取 8 组实验样本作为模型测试集, 剩余 27 组作为模型训练集。以经过处理的光谱特征峰强度和对应的血清蛋白含量分别作为模型的输入值及输出值, 建立 IABC-SVR, ABC-SVR 和 BP 三种算法的定量模型, 对测试集血清蛋白总量进行预测。最后通过均方差(MSE), 相关系数( $r$ )与建模时间分别进行对比, 结果表明通过 IABC-SVR 建立的山羊血清蛋白定量矫正模型效果最佳, 模型的相关系数为 0.990 27, 均方误差为 0.244 3, 建模时间为 1.9 s, 预测值方差均小于  $0.001\text{ g}\cdot\text{mL}^{-1}$ , 预测准确率 99.8%。实验结果表明, 应用激光拉曼光谱技术结合 IABC-SVR 算法, 对快速定量检测山羊血清蛋白含量, 具有较高的准确率和稳定性。

**关键词** 激光拉曼光谱; 血清蛋白; 改进人工蜂群优化; 支持向量机回归; 定量分析

**中图分类号:** TN247 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)02-0540-06

## 引 言

血清总蛋白(total protein, TP)是血清中各种蛋白质的总称, 是各种蛋白的复杂混合物<sup>[1]</sup>, 包含白蛋白、 $\alpha 1$  球蛋白、 $\alpha 2$  球蛋白、 $\beta$  球蛋白等, 是人体保持健康的重要组成部分。人体生理机能的变化会引起血浆中蛋白质发生质和量的变化, 检测血清生化成分, 可作为协助诊断肿瘤、肝病、肾病综合征等疾病的方法<sup>[2]</sup>。例如, 郭丽等通过对比健康人和患者的血清拉曼光谱, 为乳腺癌的诊治提供了新的方法。王玉等<sup>[3]</sup>利用不同批次人血白蛋白的拉曼光谱, 结合拉曼光谱峰位移和峰强度的相似度计算结果, 来判断血液白蛋白样品的真假伪劣。本文提出了一种基于拉曼光谱和改进人工蜂群算法优化支持向量机回归(IABC-SVR)算法快速定量检测山

羊血清蛋白含量的方法。

血清总蛋白测定方法有很多, 最主要的有凯氏定氮法、双缩脲法、酚试剂法等, 目前检测血清蛋白总量的常规方法为双缩脲法<sup>[4]</sup>, 其原理是蛋白质分子中的肽键( $-\text{CO}-\text{NH}-$ )与双缩脲试剂反应, 形成紫色化合物, 其紫色深浅程度与样品中蛋白质浓度成正比, 根据所测样品吸光度值, 在标准曲线上查出相应蛋白质含量, 从而计算出样品溶液蛋白质含量。该方法需消耗多种试剂, 步骤较繁琐, 存在污染试剂、精度差、容易交叉污染等缺点。而拉曼光谱技术是一种非破坏性测试技术, 几乎无需试样制备, 用很少量的试样就能获得足够的信号, 所表征的是分子振动的信息, 水的拉曼散射极微弱, 因而水溶液样品可直接进行测量, 这对生物大分子的研究非常有利。从拉曼光谱的谱图中可以得到频率、强度、偏振特性、峰形等信息, 在生物、材料、环保、地质等

收稿日期: 2019-12-29, 修订日期: 2020-04-27

基金项目: 国家自然科学基金项目(61675176, 61575170), 燕山大学博士基金项目(B771), 燕山大学“新锐工程”人才支持计划项目资助

作者简介: 付兴虎, 1981年生, 燕山大学信息科学与工程学院教授 e-mail: fuxinghu@ysu.edu.cn

领域得到应用。这种生化拉曼光谱技术具有灵敏度高、无需样品预处理、操作简便、信息含量丰富、能够获取分子的结构信息等特点。本文将拉曼光谱技术与 IABC-SVR 算法相结合,对山羊血清蛋白总量进行快速定量分析。

## 1 模型原理和方法

### 1.1 支持向量机回归

支持向量机分类(support vector machine, SVM)于 1995 年由 Vapnik 提出,支持向量机回归(support vector regression, SVR)是在 SVM 的基础上发展起来的,是将低维空间的非线性问题通过核函数映射到高维空间,使其在高维空间内成为一个线性问题,从而使问题的求解得到了简化<sup>[5]</sup>。设原有空间训练样本 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), (i=1, 2, \dots, l), x_i (x_i \in R^d)$ 是第  $i$  个训练集样本的输入列向量,  $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T, y_i \in R$  为对应的输出值。利用非线性映射函数  $\Phi(x)$  变换将  $l$  个训练样本映射到高维特征空间,基于结构风险最小化的原则建立高维空间线性回归函数为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \quad (1)$$

式(1)中,  $b$  为阈值;  $\alpha_i$  和  $\alpha_i^*$  为拉格朗日乘子;  $K(x_i, x_j)$  为核函数,本文选用径向基函数作为核函数。

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

约束条件为  $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C, C$  为惩罚参数。

### 1.2 改进的人工蜂群算法

标准人工蜂群算法(artificial bee colony, ABC)是由土耳其学者 Karaboga 于 2005 年提出模拟蜜蜂寻觅花蜜过程的一种群体智能算法<sup>[6]</sup>。在人工蜂群的框架下,蜜源的位置表示要优化问题的可能候选解,蜜源中花蜜的数量与候选解的适应度(目标值函数)相对应,而蜂群的规模等于待优化解的数量。ABC 算法更新种群采用随机选取的方法,在 ABC 算法搜索过程中存在收敛速度较慢、不能实现全局收敛的问题,针对此问题,提出改进的人工蜂群算法(improved artificial bee colony, IABC),在对蜜源更新公式中引入柯西变异算子<sup>[7]</sup>。同时改进随机解生成方法,随着迭代次数增加,步长从一个较小的数逐渐增长到一个较大数值。改进后的算法保证模型前期能够进行全面的局部搜索,且后期保持全局搜索能力,并加快收敛速度。

设初始种群含有  $NP$  个解,每个解  $x_i (i=1, 2, \dots, NP)$  为一个  $d$  维向量,对应的解的适应度值  $fit$ ,个体最大更新次数为  $G$ ,最大迭代次数为  $T$ 。具体步骤为

第 1 步:初始阶段,随机产生蜜源的初始位置,即

$$x_{ij} = x_{\min} + \log_{rt}(x_{\max} - x_{\min}) \quad (3)$$

第 2 步:初始蜜源周围搜索产生一个新的蜜源,即

$$v_{ij} = (x_{ij} + x_{ij} \times C_{ij}) + \varphi_{ij}(x_{ij} - x_{kj}) \quad (4)$$

其中,  $C_{ij}$  为对应  $x_{ij}$  服从柯西分布  $C(0, 1)$  的随机数,  $k \in \{i=1, 2, \dots, NP\}, j \in \{1, 2, \dots, d\}$  且  $k \neq i$ 。  $\varphi_{ij}$  为  $[-1,$

$1]$  内的随机数。

第 3 步:评价蜜源的适用度,如果新的蜜源适应度高于原先的蜜源,则新的蜜源代替原先的蜜源,否则保留原先的蜜源,即

$$fit_i = \begin{cases} \frac{1}{f_i + 1} & f_i \geq 0 \\ 1 + abs(f_i) & f_i < 0 \end{cases} \quad (5)$$

其中  $f_i$  为解的函数值。此外,基于轮盘赌原则,找到新蜜源被跟随的概率  $P_i$  为

$$P_i = \frac{fit_i}{\sum_{i=1}^{NP} fit_i} \quad (6)$$

某蜜源经过  $G$  次循环后没有得到优化,则放弃该蜜源寻找下一个。若运算迭代次数达到  $T$  或者最优适应度,运算结束,否则返回式(4),迭代出最优解。

## 2 实验部分

### 2.1 仪器与样品

本次实验使用的山羊血清为经过无菌过滤处理的山羊血清,血清蛋白初始浓度为  $0.042\ 34\ \text{g} \cdot \text{mL}^{-1}$ 。实验中用于稀释血清浓度的试剂为生理盐水,生理盐水对实验结果影响极小且不影响生物样本的活性。按体积比配置不同浓度的血清蛋白样本,配置的单个液体样本体积为  $3\ \text{mL}$ ,共计 35 组,随机选取 8 组实验样本作为模型测试集,剩余 27 组作为模型训练集。配置好的样本在  $4\ ^\circ\text{C}$  恒温箱中静置 24 h,确保样本混合均匀。光谱测量过程中易受外界光干扰,所以在实验中进行避光处理,防止外界干扰。

本次实验使用的是 BWS465-785S 型拉曼光谱仪,由美国必达泰克公司生产,光谱测量范围  $0 \sim 3\ 500\ \text{cm}^{-1}$ ,激发光源波长为  $785\ \text{nm}$ ,设置拉曼光谱激光功率为  $300\ \text{mW}$ ,每个样本连续扫描取光谱稳定后的数值,积分时间  $10\ 000\ \text{ms}$ ,光谱仪的采集分辨率为  $5\ \text{cm}^{-1}$ 。

### 2.2 光谱采集与数据处理

使用拉曼光谱仪采集样本拉曼光谱,通过 BWRam4 软件进行光谱读取。测量光谱前,将拉曼光谱仪打开预热  $20\ \text{min}$ 。用超纯水清洗样品池,控干池内水分,并使用擦镜纸清理样品池表面(测量各组样品前重复此步骤)。将试剂样本缓慢移入样品池,避免气泡进入,放入遮光罩中进行避光处理,待样品静置  $2\ \text{min}$  后,进行光谱采集。拉曼光谱处理过程如图 1 所示。

通过对图 1 分析可知,随着样品中血清蛋白浓度变化,拉曼光谱原始数据强度发生明显变化。原始光谱信噪比较高,影响了对特征峰的观察和数据处理。在建模之前对数据进行预处理,随机选取一条拉曼光谱为例进行处理。

每组样品采集前进行暗电流采集并扣除。采集的原始拉曼光谱如图 1(a)所示(选取光谱范围  $300 \sim 1\ 300\ \text{cm}^{-1}$ )。为了消除 CCD 噪声、采集电路的噪声以及激光功率抖动等对拉曼信号强度的影响,同时简化计算量,采用平滑处理、背景扣除、归一化三种预处理方法对测得的原始拉曼光谱进行

数据预处理。平滑采用 Savitzky-Golay 平滑法进行处理,该方法在滤除噪声的同时可保持原信号谱宽、强度及形状不发生改变,其滤波原理是局部多项式时域最小二乘法拟合<sup>[8]</sup>,处理结果如图 1(b)所示;背景扣除采用扣除基线的方法,采用非对称最小二乘平滑建立基线,可将有效拉曼信号保存下

来,去掉背景荧光信号如图 1(c)所示,黑色线代表拉曼光谱数据,红色线代表需要扣除的基线;数据归一化处理的作用是简化计算,缩小量值,提高数学预测模型的准确度,如图 1(d)所示,黑色线代表拉曼光谱扣除基线后的数据,红色线代表执行归一化处理后的数据。

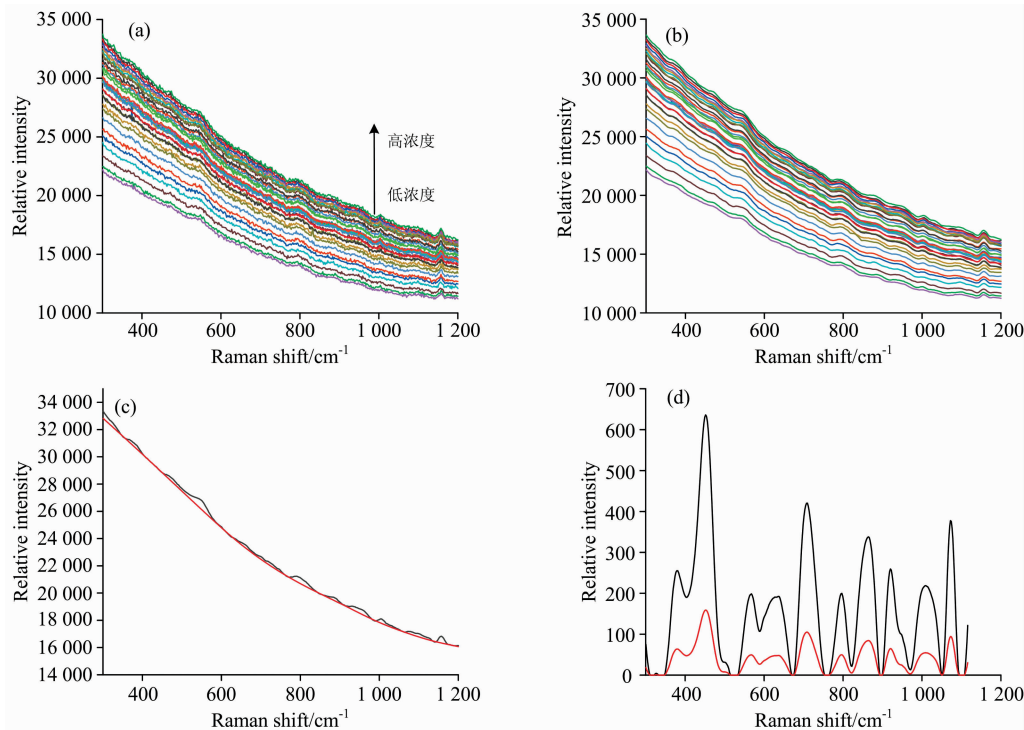


图 1 拉曼光谱处理过程

(a): 原始拉曼光谱; (b): 平滑处理; (c): 背景基线扣除; (d): 归一化处理

Fig. 1 Raman processing

(a): 26 groups of original Raman spectra; (b): Smoothing processing; (c): Background baseline deduction; (d): Normalization

### 2.3 特征峰归属

经过处理的山羊血清样本拉曼光谱,在 379.05, 451.87, 568.07, 636.12, 709.73, 797.08, 864.21, 920.15, 1 009.95 和 1 072.24  $\text{cm}^{-1}$  处表现出明显的特征峰,如图 1(d)所示。通过对比,对山羊血清蛋白特征峰进行归属<sup>[9-10]</sup>,峰值标记如图 2 所示,特征峰归属如表 1 所示。

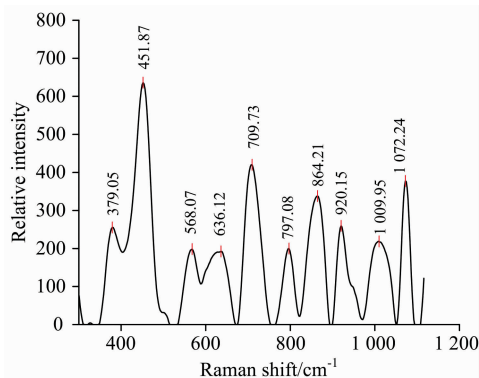


图 2 血清蛋白拉曼光谱特征峰

Fig. 2 Characteristic peak of Raman spectrum of serum protein

表 1 血清蛋白特征峰及其归属

Table 1 Characteristic peak of serum protein and its attribution

拉曼位移 / $\text{cm}^{-1}$	振动模式初步指认	主要归属
379.05	Trp	色氨酸
451.87	Man	D-甘露醇
568.07	$\nu(\text{s-s})$	二硫键伸缩振动
636.12	$\nu(\text{s-s}), \text{Tyr}$	酪氨酸
709.73	$\gamma(\text{c-s})$	辅酶
797.08	Asn	天冬酰胺
864.21	Trp	色氨酸
920.15	$\nu(\text{c-c}), \text{Man}$	D-甘露醇
1 009.95	$\nu(\text{c-n}), \text{Man}, \text{Amide-III}$	D-甘露醇, 碳氮键伸缩振动, 酰胺-III
1 072.24	Beta carotene, Tyr	$\beta$ -胡萝卜素, 酪氨酸

注:  $\nu$  伸缩振动;  $\gamma$  面外弯曲振动。

### 3 结果与讨论

在 35 组实验样品中随机选取 27 组样本作训练集,通过

拉曼光谱的数据采集及预处理,提取的特征峰强度值和训练集已知浓度信息作为输入,余下 8 组样本作为测试集,同训练集一样,将特征峰强度作为模型的输入,通过模型运算的浓度信息作为输出。分别建立基于改进人工蜂群算法优化支持向量机回归算法(IABC-SVR)的预测模型和人工蜂群算法优化支持向量机回归算法(ABC-SVR)模型,并与经典 BP(back propagation)神经网络算法建模结果比较。IABC 属于迭代优化,可优化 SVR 中的惩罚因子  $C$  和函数参数  $\sigma$ ,以提升该模型的泛化能力及预测精度。建立 IABC-SVR 模型流程如图 3 所示。

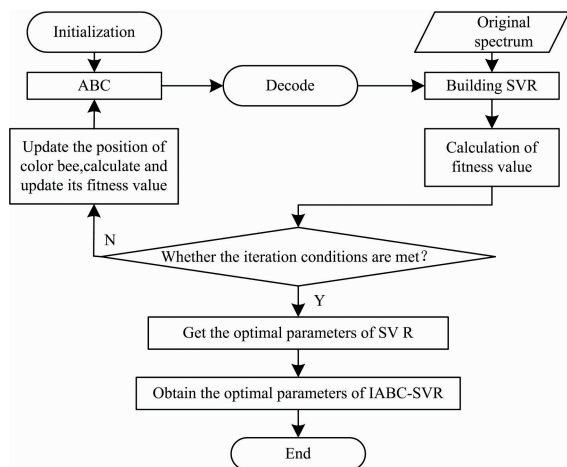


图 3 建模过程流程

Fig. 3 Modeling process flow chart

将测试集已知血清蛋白浓度与预测结果进行均方差运算来评价模型精准度,相关系数来评价模型预测关联度。样本为  $M$  的均方差定义为

$$MSE = \frac{\sum_{i=1}^M (y_i - \hat{y}_i)^2}{M} \quad (7)$$

式(7)中,  $M$  为样本个数;  $y_i$  为实际值;  $\hat{y}_i$  为预测值。

均方根误差可以显示出预测结果与真实结果平方和的平均数,由于对误差进行了平方处理,就避免了误差出现正负值而相抵消的情况,从而提高了对误差分析的准确性。相关系数  $r$  定义为

$$r = \frac{M \sum_{i=1}^M y_i \hat{y}_i - \sum_{i=1}^M y_i \times \sum_{i=1}^M \hat{y}_i}{\sqrt{\frac{1}{M} \sum_{i=1}^M y_i^2 - \left(\frac{\sum_{i=1}^M y_i}{M}\right)^2} \times \sqrt{\frac{1}{M} \sum_{i=1}^M \hat{y}_i^2 - \left(\frac{\sum_{i=1}^M \hat{y}_i}{M}\right)^2}} \quad (8)$$

$r$  值越接近 1,则反映定量预测模型精度越高,预测集建模相关度曲线如图 4 所示,相关系数为 0.990 27,显示出极高的相关性,表明模型结果较好。预测结果残差如图 5 所示,残差均小于  $0.001 \text{ g} \cdot \text{mL}^{-1}$ ,预测准确率为 99.8%。

将数据分别导入 ABC-SVR 算法和 BP 神经网络算法进行建模,并将预测模型进行对比,建模结果如图 6 所示,得到的各模型均方差(MSE),相关系数( $r$ )与建模时间(Time)如表 2 所示。

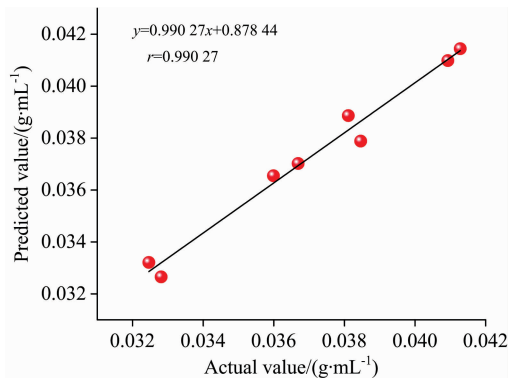


图 4 模型测试集中真实值与预测值相关度拟合  
Fig. 4 Correlation curve between actual value and predicted value in model test set

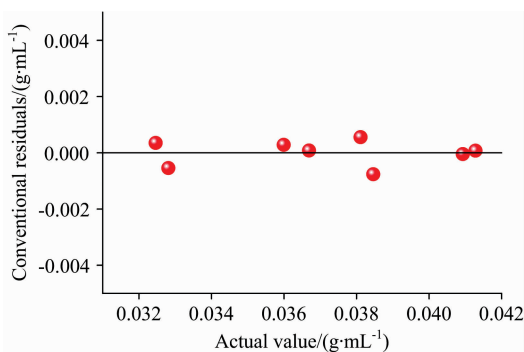


图 5 模型预测残差

Fig. 5 Residual of model prediction

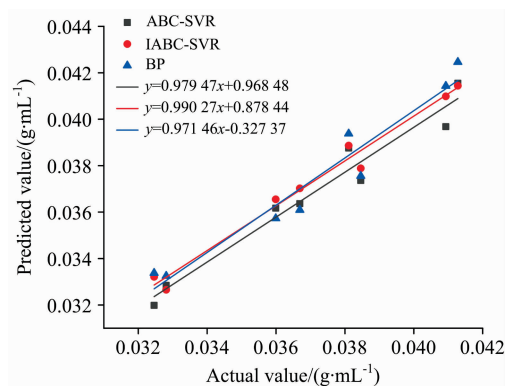


图 6 三种模型预测结果中真实值与预测值相关度曲线  
Fig. 6 Correlation curve between actual value and predicted value in three models' prediction results

表 2 三种建模方法模型评价结果

Table 2 Evaluation results of three modeling methods			
	MSE	$r$	Time/s
BP	0.681 2	0.971 46	23.5
ABC-SVR	0.321 2	0.979 47	2.5
IABC-SVR	0.244 3	0.990 27	1.9

由表 2 可知,经典 BP 神经网络算法的预测精度低于

ABC-SVR 算法建模精度, IABC-SVR 算法预测模型在 ABC-SVR 算法基础上有明显提升, 说明 IABC-SVR 算法能较好的避免局部最优解, 有较好的全局搜索能力。建模时间为 1.9 s, 具有更好的工作效率。由实验结果可以看出, 文中提出的基于 IABC-SVR 算法预测模型定量分析山羊血清蛋白总量的方法能有效实现预期功能, 并在 ABC-SVR 算法基础上提高了精度, 加快了算法收敛运算速度, 达到了更好的效果。

## 4 结 论

提出了一种基于拉曼光谱和 IABC-SVR 算法对山羊血清蛋白含量快速、无损的定量检测方法。拉曼光谱仪采集实

验样本光谱, 结合三种光谱预处理的方法对原光谱进行处理, 标定的光谱特征峰强度作为模型输入; 针对传统人工蜂群算法在区域规模较大时收敛速度逐渐减慢, 会出现效率低、精准度下降、局部最优解概率高等弊端, 改进了蜂群的初始化和位置更新方法, 旨在提高预测模型的全局搜索能力和准确性; 建立了 IABC-SVR 山羊血清蛋白定量检测模型, 对山羊血清蛋白总含量进行定量分析。结果表明, 该方法能快速、准确、无损的预测山羊血清蛋白总量, 具有较高的可行性。建立模型的相关系数为 0.990 27, 均方误差为 0.244 3, 建模时间为 1.9 s, 预测准确率为 99.8%。说明该方法能较好的完成山羊血清蛋白总量的快速定量分析。在下一步的实验中, 将样品的浓度梯度进一步细化, 增加模型的训练集和测试集, 提升模型的精度和预测能力。

## References

- [ 1 ] ZHANG Yi-yuan, TANG Hong, GUO Yan-hua, et al(张译元, 唐 红, 郭延华, 等). Xinjiang Agricultural Sciences(新疆农业科学), 2017, 54(1): 190.
- [ 2 ] CHEN Hua-cai, YANG Zhong-guo, CHEN Xing-dan, et al(陈华才, 杨仲国, 陈星旦, 等). Chinese Journal of Analysis Laboratory(分析试验室), 2005, 24(7): 17.
- [ 3 ] WANG Yu, WANG Si-huan, CHEN Jian-guo, et al(王 玉, 王思寰, 陈建国, 等). Chinese Journal of Pharmaceutical Analysis(药物分析杂志), 2012, 32(3): 537.
- [ 4 ] TANG Kai-yong, ZHOU Hong-xiang, TIAN Ya-ling, et al(唐开永, 周鸿翔, 田娅玲, 等). Food Research and Development(食品研究与开发), 2018, 39(16): 144.
- [ 5 ] WANG Chun-long, LIU Jian-guo, ZHAO Nan-jing, et al(王春龙, 刘建国, 赵南京, 等). Acta Optica Sinica(光学学报), 2013, 33(3): 0330002.
- [ 6 ] LUO Jun, LI Yan(罗 钧, 李 研). Control and Decision(控制与决策), 2010, 25(12): 1913.
- [ 7 ] GUO Yi-nan, LIU Dan-dan, CHENG Jian, et al(郭一楠, 刘丹丹, 程 健, 等). Acta Electronica Sinica(电子学报), 2011, 39(8): 1913.
- [ 8 ] ZHU Hong-yun, WANG Chang-long, WANG Jian-bin, et al(朱红运, 王长龙, 王建斌, 等). Journal of Computer Applications(计算机应用), 2015, 35(10): 3004.
- [ 9 ] GUO Jing, RONG Zhen, XIAO Rui, et al(郭 晶, 荣 振, 肖 瑞, 等). Military Medical Sciences(军事医学), 2016, 40(4): 350.
- [ 10 ] LIU Wan-hua, LIU Jian, ZHAO Xue-song, et al(刘婉华, 刘 健, 赵雪松, 等). Laser Journal(激光杂志), 2018, 39(3): 68.

# Quantitative Analysis of Goat Serum Protein Content by Raman Spectroscopy Based on IABC-SVR

FU Xing-hu, ZHAO Fei, WANG Zhen-xing, LU Xin, FU Guang-wei, JIN Wa, BI Wei-hong

School of Information Science and Engineering, The Key Laboratory for Special Fiber and Fiber Sensor of Hebei Province, Yanshan University, Qinhuangdao 066004, China

**Abstract** A method based on Raman spectroscopy and improved artificial bee colony algorithm to optimize the support vector machine regression (IABC-SVR) algorithm for rapid quantitative detection of goat serum protein content was proposed. The traditional artificial bee colony algorithm gradually slows down the convergence rate when the data area is large in size, which causes problems such as low efficiency, decreased accuracy, and high optimal local solution probability. The proposed algorithm solves these problems so that the algorithm avoids falling into the local optimal solution in the early stage of evolution, and can maintain the global search ability of the solution in the middle and late stages of evolution. Conventional methods for measuring total serum protein usually use Kjeldahl method, biuret method, etc., which has the disadvantages of slow aging and contaminated samples. This paper uses Raman spectroscopy for detection, which has the advantages of fast and non-destructive. Using goat serum as the analysis object, configure 35 groups of samples to be measured according to a certain volume ratio. Raman spectra were collected using a Raman spectrometer with a spectral collection range of  $300\sim 1300\text{ cm}^{-1}$ . Baseline correction was used to remove the fluorescent background, and Savitzky-Golay spectra were used to smooth this method smoothes the original spectrum, normalizes the spectral data, and assigns the characteristic peaks of the Raman spectrum. The experimental results show that Raman spectroscopy can characterize the information of the main chemical groups in the serum, and due to the difference in functional group concentration, the spectral characteristic peak intensity changes significantly with the concentration, so the total serum protein can be determined based on the characteristic peak information. In the experiment, based on the purchased goat serum protein content, the protein content of each group of serum samples was obtained by configuring the volume ratio of the samples. The volume of a single liquid sample was 3 mL. Eight groups of experimental samples were randomly selected as the model test set. The remaining 27 groups are used as model training sets. The processed spectral characteristic peak intensity and the corresponding serum protein content were used as the input and output values of the model to establish a quantitative model of IABC-SVR, ABC-SVR, and BP algorithms to predict the total serum protein in the test set. Finally, the mean square error (MSE), correlation coefficient ( $r$ ) was compared with the modeling time, and the results showed that the goat serum protein quantitative correction model established by IABC-SVR had the best effect. The correlation coefficient of the model was 0.990 27, and the mean square error was 0.244 3, the modeling time is 1.9 s, the variance of the predicted values are less than  $0.001\text{ g}\cdot\text{mL}^{-1}$ , and the prediction accuracy is 99.8%. The experimental results show that the laser Raman spectroscopy technology combined with the IABC-SVR algorithm has high accuracy and stability for the rapid quantitative detection of goat serum protein content.

**Keywords** Laser Raman spectroscopy; Serum protein; Improved artificial bee colony optimization; Support vector machine regression; Quantitative analysis

(Received Dec. 29, 2019; accepted Apr. 27, 2020)