基于微型近红外光谱仪的油菜籽含油率模型参数优化研究

陈斌，卢丙，陆道礼

（江苏大学食品与生物工程学院，江苏镇江 212013）

摘要：为实现油菜籽含油率快速无损检测，采用微型近红外光谱仪，结合竞争性自适应重加权(CARS)、遗传算法(GA)、连续投影算法(SPA)、无信息变量消除法(UVE)、向后区间偏最小二乘法(BIPLS)、联合区间偏最小二乘法(SIPLS)等方法优选油菜籽含油率近红外光谱特征波长，建立偏最小二乘回归(PLSR)和最小二乘支持向量机(LS-SVM)定量分析模型，同时对 LS-SVM 模型参数进行优化。研究表明，对 PLSR 模型，BIPLS+GA 优选的 26 个特征波长建模效果最好，其预测相关系数(R_p)和预测均方根误差(RMSEP)分别为 0.9330 和 0.0075，对 LS-SVM 模型，SIPLS+GA 优选的 13 个特征波长建模效果最好，预测相关系数(R_p)和预测均方根误差(RMSEP)分别 0.9192 和 0.0055。证明了波长优选和参数优化可有效简化油菜籽含油率近红外光谱定量分析模型，提高模型预测精度和稳定性，为进一步拓展微型近红外光谱仪的应用提供技术参考。

关键词：油菜籽；含油率；波长优选；最小二乘支持向量机

文章篇号：1673-9078(2015)8-286-292 DOI: 10.13982/j.mfst.1673-9078.2015.8.045

Parameter Optimization of Rapeseed Oil Content Model Using a Miniature Near-infrared Spectrometer

CHEN Bin, LU Bing, LU Dao-li

(School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: To select characteristic NIR wavelengths for rapid and nondestructive detection of rapeseed oil content, a miniature near-infrared (NIR) spectrometer was used in combination with methods including competitive adaptive reweighted sampling (CARS), genetic algorithm (GA), successive projections algorithm (SPA), uninformative variable elimination (UVE), backward interval partial least squares (BIPLS), and synergy interval partial least squares (SIPLS). Subsequently, partial least squares regression (PLSR) and least squares-supported vector machine (LS-SVM) regression model were established and the parameters of LS-SVM model were optimized. The results showed that for PLSR model, 26 characteristic wavelengths selected by BIPLS + GA produced the optimum model, where the correlation coefficient (R_p) and root mean square error of prediction (*RMSEP*) for prediction sets were 0.9330 and 0.0075, respectively. For LSS-VM model, 13 characteristic wavelengths selected by SIPLS+GA produced the optimum model and the correlation coefficient (R_p) and root mean square error of prediction (*RMSEP*) for prediction sets were 0.9192 and 0.0055, respectively. The above results demonstrate that parameter optimization and wavelength selection can not only effectively simplify the quantitative analysis model for rapeseed oil content based on miniature NIR spectrometry, but also enhance prediction accuracy and prediction stability. These data provide useful technical references for further applications of miniature NIR spectrometry.

Key words: rapeseed; oil content; wavelength selection; least squares support vector machine

油菜是我国 5 个种植面积超亿亩的主要农作物（水稻、玉米、小麦、大豆、油菜）之一在农村经济和人民生活中占有重要位置。油菜籽含油量约 40%，是我国最重要的食用植物油源，油菜籽含油率的常规测定方法为核磁共振及索氏抽提法，但是这些测定方法会存在样品需要前处理、操作复杂、周期长及成本高等问题，而且所用化学试剂还有可能对环境造成污染，这些情况并不适合我国油菜籽分散种植和大批量样品检测的要求。

近红外光谱分析技术(Near infrared spectroscopy, NIRS)是 20 世纪 80 年代以来发展最快的高效、快速、

收稿日期：2014-11-26

基金项目：国家自然科学基金项目（31171697）；国家重大科学仪器开发专项(2014YQ491015)；江苏高校优势学科建设工程资助项目

作者简介：陈斌（1960-），男，博士，教授，主要从事近红外光谱分析和农产品无损检测技术的研究

跨学科现代分析技术，国外很早就把它应用在了油菜籽的快速无损检测中，Daul 等[1]研究了三种不同近红外光谱仪的油菜籽含油率分析模型，Hom 等[2]研究了小样品油菜籽含油率近红外光谱分析模型，Petisco 等[3]研究了不同品种油菜籽的可见及近红外光谱分析模型，Sidhu 等[4]研究了二极管阵列检测型近红外光谱仪的油菜籽含油率分析模型。国内方面高建芹等[5]研究了小样品油菜籽含油率及脂肪酸分析模型，智文良等[6]研究了一种国产近红外光谱仪在油菜籽成分分析中的应用，康月琼等[7]研究了特定地区油菜籽常见成分的近红外光谱分析模型。这些研究都建立了稳定可靠的模型。然而，这些模型大都是基于傅里叶变换型、光栅扫描型和阵列检测型近红外光谱仪而建立的，它们要么存在较多精密的移动部件要么扫描速度较慢要么价格比较昂贵，因此在农产品快速无损检测的普及应用中存在一定难度。

本研究采用基于线性渐变滤光片(Linear variable filter，LVF)分光原理的微型近红外光谱仪，是目前市场上最小的近红外光谱仪之一，与上述和常见的基于傅里叶变换和微光机电系统(Micro optic electro mechanical systems，MOEMS)的微型近红外光谱仪相比，具有紧凑、轻便和无移动部件等特点[8]，就通用性、便携性及价格而言也可以满足农产品快速无损检测的要求，同时很容易进行二次开发，但该光谱仪受线性阵列铟砷化镓(InGaAs)检测器规格的制约，采集的光谱仅有 125 个变量，因此应用开发过程中对光谱的高效处理也变得更加重要。目前对这种类型的光谱仪进行系统模型优化的研究也较少。本文在光谱预处理和波长优选的基础上对模型参数进一步优化，以进一步提高模型稳定性和预测性能，为基于微型近红外光谱仪的油菜籽含油率快速定量分析提供参考。

1 材料与方法

1.1 实验材料

油菜籽样品于 2014 年 5 月份采集于镇江丹阳、句容两地的农田，品种为甘蓝型双低油菜秦优 7 号及杨油 6 号，挑选正常生长的油菜植株，每棵植株按其自身高度分为上部、中部和下部分开采集并编号，室内阴干后人工脱壳并过筛去杂，检测前保存于 4 ℃左右冰箱，近红外光谱与含油率测定前恢复至室温 25 ℃，空气相对湿度 40%，取其中 85 份作为校正集，37 份作为预测集。

1.2 仪器与光谱采集

试验使用美国 JDSU 公司生产的 MicroNIR-1700 光谱仪(图 1)，它在 Φ45×42 mm 的体积上集光源、滤光片和检测器等于一体，不需要任何移动部件，重量仅为 60 g 左右。具体参数：光源使用双集成真空钨灯，分光元件为线性渐变滤光片，探测器采用 128 线元非制冷铟镓砷（InGaAs）二极管阵列检测器，工作波长范围 950~1650 nm，运行环境-20 ℃~40 ℃，光谱分辨率 12.5 nm，积分时间 11 ms，扫描次数 25 次。把油菜籽样品除杂后均匀装在方形样品杯中，连续采集不同部位的 3 次漫反射光谱，平均后作为最终光谱。

图 1 MicroNIR-1700 光谱仪

Fig.1 MicroNIR-1700 spectrometer

1.3 含油率的测定

将采集光谱后的样品烘干并粉碎后采用索氏残余留量法测定样品含油率[9]。校正集和预测集样品的划分采用 Kennard-Stone 法。其实际含油率分布如表 1，可以看出校正集的含油率范围大于预测集的范围，这样有利于保证预测模型的稳健性。

表 1 120 个样品的含油率

Table 1 Oil content in 120 samples

Data set	Samples	Min	Max	Mean	S.D.
Calibration	85	0.3800	0.4831	0.4301	0.4301
Prediction	37	0.3931	0.4496	0.4496	0.4206

1.4 模型的建立及评价

利用 Matlab7.8.0(Mathworks,USA)和江苏大学近红外工作室 NIRSA 数据处理系统 Ver4.3.1 完成近红外光谱预处理、波长优选、模型的建立及参数优化。分别建立偏最小二乘回归(Partial least squares regression, PLSR)和最小二乘支持向量机(Least square support vector machine, LS-SVM)定量分析模型，选取校正相关系数(Correlation coefficient of calibration, R_c)，预测相关系数(Correlation coefficient of prediction, R_p)，校正均方根误差(Root mean square error of calibration, RMSEC)，预测均方根误差(Root mean square error of prediction, RMSEP)作为模型的评价标准，一个好的模型应该有较高的 R_c 和 R_p，较低的 RMSEC 和 EMSEP。

R_p 越高且 RMSEP 越小，模型预的测能力越强。RMSEC 和 RMSEP 越接近，模型的预测稳定性就越好[10]

2 结果与讨论

2.1 光谱预处理

原始光谱由于各种因素的影响，包含了除油菜籽自身近红外光谱信息之外的其它噪声信息，因此有必要加以预处理。对于采集到的 960 nm 到 1650 nm 范围内的数据进行 5 点移动平滑滤波(Moving average filter, MAF)、5 点卷积平滑(Savitzky golay filter, SGF)、多元散射校正(Multiplicative scatter correction, MSC)、归一化(Normalization)、中心化(Autoscaling)、标准正态变换(Standard normal variable transformation, SNV)+去趋势、小波阈值去噪等处理，然后建立全光谱 PLSR 和 LS-SVM 模型，建模效果较好的预处理方法如表 2，比较后确定 PLSR 和 LS-SVM 模型的最佳光谱预处理方法分别为小波阈值去噪和 SNV+去趋势。

2.2 基于 PLSR 建模的特征波长优选

PLSR 能够在自变量存在严重多重相关性和样本点数少于变量个数的条件下进行回归建模；它将不再直接考虑因变量与自变量的建模关系，而是对变量系统中的信息进行综合筛选，从中选取若干对系统具有最佳解释能力的新综合变量进行回归建模。这个过程排除了对因变量无解释作用的噪声。因此与普通最小二乘回归相比，PLSR 模型更具先进性，其计算结果也更可靠。

表 2 不同预处理方法所建模型校正和预测结果(PLSR 模型和 LS-SVM 模型)

Table 2 Result of calibration and prediction for the models established using different pretreatment methods

Method		Calibration		Prediction	
		R_c	RMSEC	R_p	RMSEP
PLSR	Raw	0.9737	0.0040	0.9150	0.0086
	MAF	0.9563	0.0052	0.9121	0.0084
	SGF	0.9626	0.0048	0.9043	0.0089
	中心化	0.9737	0.0040	0.9150	0.0086
	归一化	0.9706	0.0044	0.8874	0.0080
	小波阈值	0.9672	0.0045	0.9169	0.0084
LS-SVM	Raw	0.9654	0.0051	0.6946	0.0092
	MAF	0.9680	0.0049	0.7156	0.0090
	SGF	0.9784	0.0040	0.7411	0.0087
	MSC	0.9965	0.0016	0.7094	0.0098
	SNV+去趋势	0.9496	0.0061	0.8452	0.0074
	归一化	0.9763	0.0043	0.7727	0.0083

表 3 不同变量优选方法所建 PLSR 模型校正和预测结果

Table 3 Result of calibration and prediction for the PLSR model established using different variable selection methods

Method	No. of variables	No. of factors	Calibration		Prediction	
			R_c	RMSEC	R_p	RMSEP
None	125	15	0.9672	0.0040	0.9168	0.0084
CARS	20	15	0.9538	0.0054	0.9243	0.0077
GA	27	10	0.9402	0.0057	0.9226	0.0083
SPA	14	14	0.9396	0.0061	0.9263	0.0083
BIPLS	51	10	0.9433	0.0059	0.9324	0.0075
SIPLS	41	10	0.9396	0.0061	0.9247	0.0067
BIPLS+GA	26	9	0.9418	0.0060	0.9330	0.0075
SIPLS+GA	27	9	0.9419	0.0060	0.9215	0.0075

对于 PLSR 定量分析模型而言，常用的波长优选方法主要有：竞争性自适应重加权法(Competitive adaptive reweighted sampling, CARS)、遗传算法(Genetic algorithms, GA)、连续投影算法(Successive projections algorithm, SPA)、向后区间偏最小二乘(Back interval PLS, BIPLS)、联合区间偏最小二乘(Synergy interval PLS, SIPLS)、向后区间偏最小二乘(BIPLS)+遗传算法(GA)、联合区间偏最小二乘(SIPLS)+遗传算法(GA)等。采用上述几种波长优选方法后较好的建模结果如表 3。可以看出较好的波长优选方法为 SPA、BIPLS+GA 和 CARS，其优选结果分别对应图 2、图 3 结合图 4、图 5。其中 BIPLS+GA 优选的 26 个变量建立的模型可以把预测相关系数(R_p)从 0.9168 提高的 0.9330，把预测均方根误差(RMSEP)从 0.0084 降到 0.0075。建模因子从 15 个减少到 9 个，有效简化了建模复杂性和冗余性，提高了模型预测能力和稳定性，比较后确定 PLSR 定量分析模型最优波长优选方法为 BIPLS+GA。与 Daul 等[1](R_p=0.979,

SEP=0.43)和 Petisco 等[3](R_p=0.98, SEP=0.62)的研究结果相比而言，由于其所用近红外光谱仪有效采集波长为 400 nm~2500 nm，波长变量数是本文的 9 倍左右，波长分辨率和精度也较高，所以预测结果也更好。Sidhu 等[4](R_p=0.84, SEP=0.61)采用的是二极管阵列检测器近红外光谱仪，光谱采集范围(950~1650)和波长点数都与本研究类似，但由于研究对象为小量油菜籽样品，需要特定的样品杯，还有建模方法的不同，所以预测相关系数 R_p 为 0.84 比本实验结果稍差。

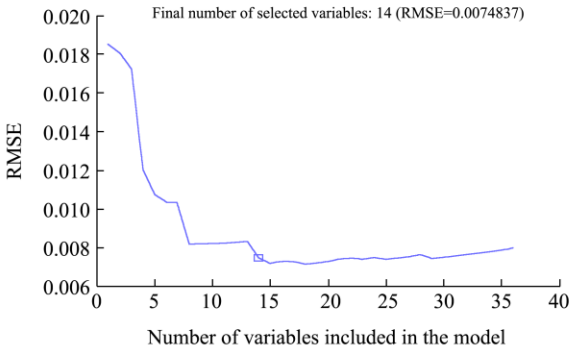

图 2 连续投影法优选波长

Fig.2 Plot of variable selection by SPA

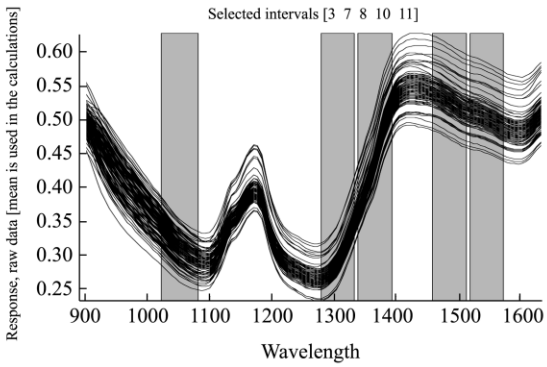

图 3 向后区间偏最小二乘法优选波长

Fig.3 Plot of variable selection by BIPLS

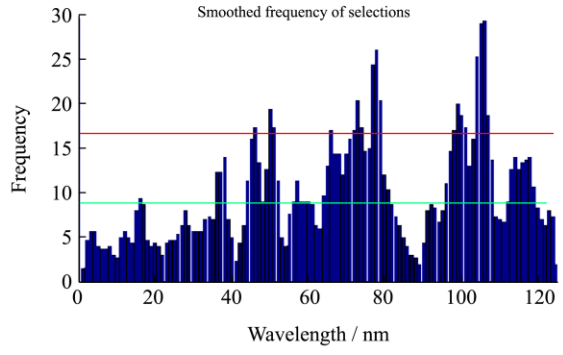

图 4 遗传算法优选波长

Fig.4 Plot of variable selection by GA

2.3 基于 LS-SVM 建模的波长优选

向量机(Support vector machine ,SVM)是数据挖掘技术中的一种常用方法，可以高效率的处理回归问题，在解决小样本、非线性及高维模式识别等问题中具有很大优势，LS-SVM 是 SVM 的改进，它采用最小方差损失函数，并用不等式约束替代原来的等式约束，将繁琐的二次规划问题转化为线性方程组问题，有效的提高了大样本学习的求解速率，鲁棒性较好，同时所需优化的参数较少，广泛的应用于回归分析。

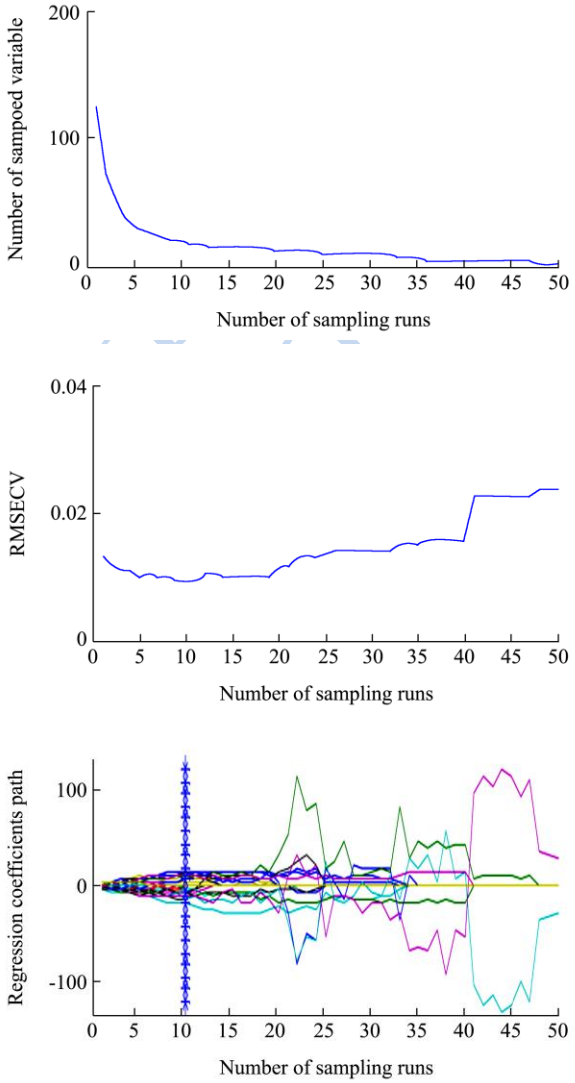

图 5 竞争性自适应重加权算法优选波长

Fig.5 Plot of variable selection by CARS

本文为了加快建模学习效率和有效提取光谱信息采用了竞争性自适应重加权法(CARS)、无信息变量消除法(Uninformative variable selection, UVE)、向后区间偏最小二乘(BIPLS)、联合区间偏最小二乘(SIPLS)、向后区间偏最小二乘(BIPLS)+竞争性自适应重加权(CARS)、联合区间偏最小二乘(SIPLS)+遗传算法(GA)等方法筛选 LSSVM 模型的特征输入变量，并选取径向基函数(Radial basis function, RBF)作为模型的核函数，其中核函数参数σ^2 和正则化参数γ首先使用常用

的网格搜索法结合留一交互验证法进行初步确定。由表 4 可以看出经波长优选后的 LS-SVM 模型的预测能力和稳定性能有很大提高，建模效果较好的优选方法为 SIPLS+GA 和 UVE，其筛选结果图分别对应图 6 结合图 7、图 8。其中 SIPLS+GA 方法优选的 13 个变量建立的最小二乘支持向量机(LS-SVM)定量分析模型可以把预测相关系数(R_p)从 0.8452 提高到 0.9175，把预测均方根误差(RMSEP)从 0.0074 降到 0.0056，有效的降低了建模复杂度，提高了模型预测能力和稳定性。与 PLSR(R_p=0.9330,RMSEP=0.0075)建模结果比较可以发现，虽然预测相关系数(R_p)稍低，为 0.9175 对比 0.9330，但预测均方根误差(RMSEP)为 0.0056 对比 0.0075，模型更加稳定，同时 LS-SVM 建模时间也更短，综合比较表 4 选取 SIPLS+GA 优选方法。建模结果和国内高建芹等[5](R_p=0.97)，智文良等[6](R_p=0.9517)分别基于傅里叶变换近红外光谱仪和光栅扫描型光谱仪建立的菜籽含油率定量分析模型相比还有一定差距，但考虑到本光谱仪的便携性及较少的波长变量数，已经达到了较好的建模效果，并且与康月琼等[7](R_p=0.9338)基于傅里叶变换光谱仪的研究结果已经十分接近。

表 4 不同变量优选方法所建 LS-SVM 模型校正和预测结果

Table 4 Result of calibration and prediction for LS-SVM model established using different variable selection methods

Method	No. of variables	γ	σ^2	Calibration		Prediction	
				R_c	RMSEC	R_p	RMSEP
None	125	1.2247×10^5	7.4948×10^4	0.9496	0.0061	0.8452	0.0074
CARS	24	5.4611×10^4	3.3806×10^6	0.9663	0.0050	0.8762	0.0069
UVE	38	2.5073×10^4	1.7009×10^5	0.9330	0.0068	0.9046	0.0063
BIPLS	82	3.2622×10^4	1.9872×10^5	0.9570	0.0055	0.8406	0.0083
SIPLS	41	1.0231×10^3	2.6193×10^4	0.9599	0.0052	0.8712	0.0076
BIPLS+CARS	22	2.8706×10^4	9.8456×10^5	0.9429	0.0062	0.8858	0.0073
SIPLS+GA	13	7.7814×10^2	2.2888×10^4	0.9229	0.0072	0.9175	0.0056

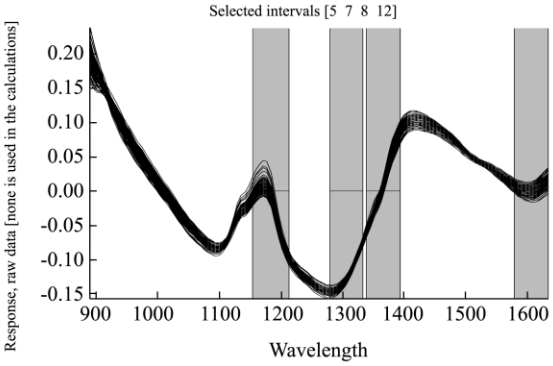

图 6 联合区间偏最小二乘筛选波长

Fig.6 Plot of variable selection by SIPLS

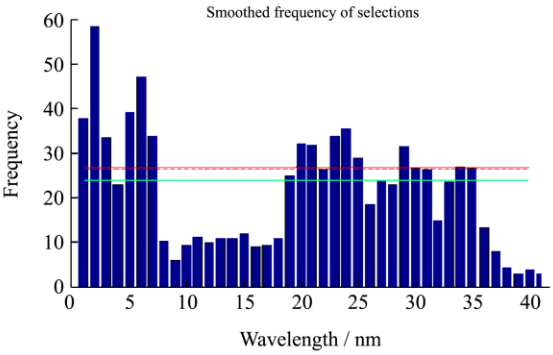

图 7 遗传算法优选波长

Fig.7 Plot of variable selection by GA

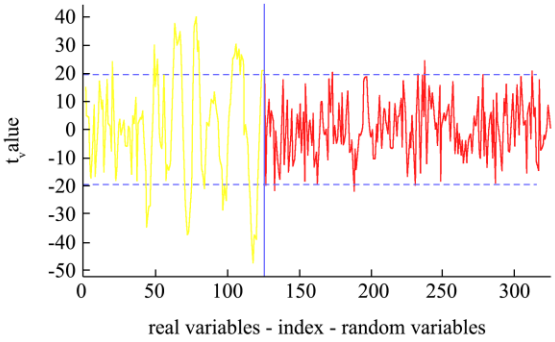

图 8 无信息变量消除法优选波

Fig.8 Plot of variable selection by UVE

2.4 LS-SVM 模型的参数优化

在 LS-SVM 回归建模中，本文选取的是泛化能力较强的径向基(RBF)核函数作为 LS-SVM 的核函数，而其中正则化参数γ和核函数参数σ^2 在很大程度上决定了 LS-SVM 的学习能力和预测能力，因此需要寻找到γ和σ^2 的最优组合。目前，对于这两个参数的选择方法一般是网格搜索法结合留一交互验证法，通过不断的实验来完成，效率较低。而群体智能优化算法作为一类新型的进化算法，对目标函数的特性没有要求，且可以再较短的时间内求解出传统方法解决不了的大

规模复杂问题的最优解（或次优解），因此得到了广泛的应用和认同，逐渐成为现代优化算法领域的研究热点，本文选取了较为新颖的粒子群优化算法[11](Particle swarm optimization, PSO)、布谷鸟算法[12](Cuckoo search，CS)和萤火虫算法[13](Firefly Algorithm, FA)作为 LSSVM 模型参数的优化方法。

PSO 优化算法由 Kennedy 博士和 Eberhart 博士在 1995 年提出的基于群体智能进化计算技术。其基本思想是根据个体与群体之间的信息传递及信息共享来寻找最优解。算法初始化为一群随机粒子，然后通过迭代寻找都最优解，与常见的遗传算法相比没有交叉和变异操作，具有容易实现、精度高和收敛速度快等特点。本文中该算法初始化过程时设定学习因子 c_1=c_2=2，最大进化次数 200，种群规模 24。FA 是由剑桥学者 Yang 在 2007 年提出的一种模拟自然界萤火虫信息交流行为的随机优化算法，具有参数设置简单、寻优精度高和全局优化能力强等特点。本文中该算法初始化时设定步长因子α=1，最大吸引度β=1.0，萤火虫数目为 15，迭代次数 500。CS 算法[13]是由剑桥大学 Yang 和拉曼工程大学的 Suash Deb 在 2009 年提出的一种模拟布谷鸟寻巢产卵行为的智能优化算法，具有参数设置少、收敛速度快和全局搜索能力强等特点，研究表明，布谷鸟搜索算法的效率远大于粒子群优化算法和遗传算法，本文采用的是基于莱维飞行规则的布谷鸟算法，算法初始化时设定鸟巢数目为 15，步长α=1，宿主鸟发现外来鸟蛋的概率 P=0.25，最大迭代次数 1500。

表 5 不同参数优化方法所建 LS-SVM 模型校正和预测结果

Table 5 Result of calibration and prediction for LS-SVM model established using different parameter optimization methods

Method	No. of variables	γ	σ^2	Calibration		Prediction	
				R_c	RMSEC	R_p	RMSEP
UVE+LSSVM+网格法	38	2.5073×10^4	1.7009×10^5	0.9330	0.0068	0.9046	0.0063
UVE+LSSVM+PSO	38	3.5947×10^4	3.8044×10^3	0.9399	0.0027	0.9134	0.0060
UVE+LSSVM+FA	38	1.0576×10^5	5.9652×10^3	0.9434	0.0062	0.9142	0.0060
UVE+LSSVM+CS	38	4.4736×10^4	3.2628×10^3	0.9439	0.0062	0.9147	0.0060
BIPLS+CARS+LSSVM+网格法	22	2.8706×10^4	9.8456×10^5	0.9429	0.0062	0.8858	0.0073
BIPLS+CARS+LSSVM+PSO	22	3.7541×10^5	6.7735×10^3	0.9490	0.0024	0.8849	0.0074
BIPLS+CARS+LSSVM+FA	22	3.6873×10^5	1.2076×10^4	0.9432	0.0062	0.8865	0.0073
BIPLS+CARS+LSSVM+CS	22	3.0820×10^5	9.2409×10^3	0.9443	0.0061	0.8866	0.0073
SIPLS+GA+LSSVM+网格法	13	7.7814×10^2	2.2888×10^4	0.9229	0.0072	0.9175	0.0056
SIPLS+GA+LSSVM+PSO	13	7.9196×10^3	1.9952×10^2	0.9361	0.0069	0.9192	0.0055
SIPLS+GA+LSSVM+FA	13	1.4101×10^5	1.5913×10^3	0.9317	0.0071	0.9177	0.0055
SIPLS+GA+LSSVM+CS	13	5.0126×10^4	1.0000×10^3	0.9311	0.0071	0.9187	0.0055

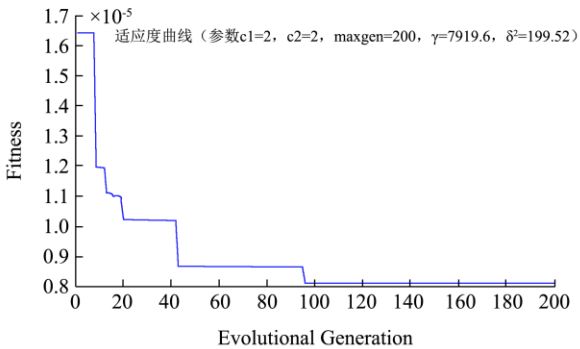

图 9 PSO 优化过程适应度曲线

Fig.9 Fitness curve of PSO optimization

表 4 表明波长优选有效提高了 LSSVM 模型的预测能力和稳定性，其中较好的特征波长优选方法有 UVE、BIPLS+CARS 和 SIPLS+GA，因此本文针对经过这三种方法优选后建立的 LS-SVM 模型进行参数优化。优化过程中以预测集的预测均方根误差作为适应度函数[14]，通过寻找最小的预测均方根误差而达到寻找最佳预测模型的目的，参数优化后具体建模结果如表 5，可以发现经过参数优化的最小二乘支持向量机(LS-SVM)模型预测能力和稳定性都有一定程度的提高。对于 UVE 优选的特征变量，最优参数优化方法是布谷鸟搜索算法(CS)；对于 BIPLS+CARS 优选的特征变量，最优参数优化算法也是布谷鸟算法；对于 SIPLS+GA 优选的特征变量，最优参数优化算法是粒子群算法(PSO)。其中经过 UVE 优选的 38 个变量所建模型优化效果最为明显，预测相关系数(R_p)从 0.9046 提高到 0.9147，预测均方根误差(RMSEP)从 0.0063 降到 0.0060；SIPLS+GA 优选的变量次之，预测相关系数(R_p)从 0.9175 提高到 0.9192，预测均方根误差(RMSEP)从 0.0056 降到 0.0055；经 BIPLS+CARS

优选的变量所建模型变化不明显，预测相关系数(R_p)从 0.8858 提高到 0.8866，预测均方根误差(RMSEP)基本不变。综合比较后，LS-SVM 定量分析模型的有效波长优选方法选取 SIPLS+GA，有效模型参数优化方法为PSO。其中 PSO 的参数优化过程及最终预测结果分别对应图 9 和图 10。

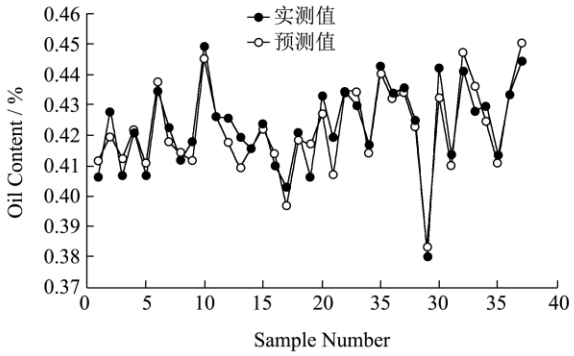

图 10 SIPLS+GA+PSO 参数优化后模型预测结果

Fig.10 Prediction results of the model after SIPLS + GA + PSO optimization

3 结论

3.1 建立了基于线性渐变滤光片型微型近红外光谱仪的油菜籽含油率定量分析模型，针对 PLSR 和 LS-SVM 两种不同建模方法采取了不同的光谱预处理和波长优选方法，并建立了全光谱和优选后特征波长的定量分析模型，同时对 LS-SVM 模型的两个重要参数进行了优化。比较了不同波长优选方法所优选特征波长的建模预测效果，其中 PLSR 定量分析模型中 BIPLS+GA 优选方法最好，其预测相关系数(R_p)和预测均方根误差(RMSEP)分别是 0.9330 和 0.0075；LS-SVM 定量分析模型中 SIPLS+GA 优化方法最好，再经过参数优化后模型预测能力和稳定性都有一定程度提高，预测相关系数(R_p)和预测均方根误差(RMSEP)分别是 0.9192 和 0.0055。结果表明小波阈值消噪+BIPLS+CARS-PLSR 和 SNV+去趋势+SIPLS+GA-LS-SVM 定量分析模型能够有效减少建模变量数，提高模型预测能力和稳定性，体现出波长优选及模型参数优化对于提高模型性能的有效性，同时发现优化后的 PLSR 模型预测相关系数稍高，LS-SVM 模型预测均方根误差更低，建模时可根据需要进行选取。

3.2 与前人基于傅里叶变换型、光栅扫描型和阵列检测型等通用型近红外光谱仪所建模型相比，本文所建模型预测能力稍低，这一方面与这些类型近红外光谱仪有较宽的波长扫描范围和较高的分辨率有关，另一方面也与建模样本数目较大有关，但本文基于线性渐变滤光片型微型近红外光谱仪的菜籽含油率定量分析模型也取得了很好的效果，比较而言仪器也更加紧凑、便携和易于校准，还可以根据具体需求结合手机、平板等智能设备实现二次开发，同时与基于傅里叶变换型和微光机电系统(MOMES)的微型近红外光谱仪相比也具有无移动部件、便于维护和成本低的特点。本文的不足是仅使用了两个品种的 122 个不同样本，后续研究会继续扩大样本范围，并在建模时软件中集成温度补偿程序，进一步优化模型。

参考文献

[1] Daun J K, Clear K M, Williams P. Comparison of three whole seed near-infrared analyzers for measuring quality components of canola seed [J]. Journal of the American Oil Chemists' Society, 1994, 71(10): 1063-1068

[2] Hom N H, Becker H C, Möllers C. Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds [J]. Euphytica, 2007, 153(1-2): 27-34

[3] Petisco C, Garcia-Criado B, Vázquez-de-Aldana B R, et al. Measurement of quality parameters in intact seeds of Brassica species using visible and near-infrared spectroscopy [J]. Industrial Crops and Products, 2010, 32(2): 139-146

[4] Sidhu H K, Haagenson D M, Rahman M, et al. Diode array near infrared spectrometer calibrations for composition analysis of single plant canola (brassica napus) seed. 2014, 30(1):69-76

[5] 高建芹,张洁夫,浦惠明,等.近红外光谱法在测定油菜籽含油量及脂肪酸组成中的应用[J].江苏农业学报,2007, 23(3):189-195
GAO Jian-qin, ZHANG Jie-fu, PU Hui-ming, et al. Analysis of oil oleic and erucic acid contents in rapeseed by near infrared reflectance spectroscopy (NIRS) [J]. Jiangsu Journal of Agricultural Sciences, 2007, 23(3): 189-195

[6] 智文良,信晓阳,崔建民,等.一种国产近红外仪分析油菜籽三种品质参数[J].中国油料作物学报,2012,34(3):305-310
ZHI Wen-liang, XIN Xiao-yang, CUI Jian-min, et al. Determination of three major quality parameters of rapeseed with near infrared analyzer NYDL-3000 [J]. Chinese Journal of Oil Crop Sciences, 2012, 34(3): 305-310

[7] 康月琼,郝风,柴勇,等.油菜品质近红外检测模型建立的研究[J].中国农学通报,2011,27 (5):144-148.
KANG Yue-qiong, HAO Feng, CHAI Yong, et al. Study on construction of determination of rapeseed quality with near infrared spectroscopy [J]. Chinese Agricultural Science Bulletin, 2011, 27(5): 144-148

[8] Lutz O, Bonn G K, Rode B M, et al. Reproducible quantification of ethanol in gasoline via a customized mobile near-infrared spectrometer [J]. Analytica Chimica Acta, 2014, 826: 61-68

[9] NY/T 1285-2007,油料种籽含量的测定残余法[S]
NY/T 1285-2007, Determination of oil content in oilseeds by residue methods [S]

[10] 陆婉珍,袁洪福,褚小立.近红外光谱仪器[M].北京:化学工业出版社,2010
LU Wan-zhen, YUAN Hong-fu, CHU Xiao-li. Near infrared spectrometer instrument [M]. Bei jing: Chemical Industry Press, 2010

[11] Chamkalani A, Zendehboudi S, Bahadori A, et al. Integration of LSSVM technique with PSO to determine asphaltene deposition [J]. Journal of Petroleum Science and Engineering, 2014, 124: 243-253

[12] Bhandari A K, Soni V, Kumar A, et al. Cuckoo search algorithm based satellite image contrast and brightness enhancement using DWT-SVD [J]. ISA Transactions, 2014, 53(4): 1286-1296

[13] Mahapatra S, Panda S, Swain S C. A hybrid firefly algorithm and pattern search technique for SSSC based power oscillation damping controller design [J]. Ain Shams Engineering Journal, 2014, 5(4): 1177-1188

[14] Long B, Xian W, Li M, et al. Improved diagnostics for the incipient faults in analog circuits using LSSVM based on PSO algorithm with mahalanobis distance [J]. Neurocomputing, 2014, 133: 237-248