

基于改进蝙蝠算法的红外光谱特征选择

陈媛媛^{1,3}, 王志斌^{1,2,3}, 王召巴^{1,2}

- (1. 电子测试技术重点实验室, 山西 太原 030051;
2. 仪器科学与动态测试教育部重点实验室, 山西 太原 030051;
3. 山西省光电信息与仪器工程技术研究中心, 山西 太原 030051)

摘要: 特征选择是红外光谱定性与定量分析中的重要环节之一。为了解决传统特征选择方法可调参数多、收敛速度慢、精度低、易早熟等不足,对基本蝙蝠算法进行了离散化改进以适用于离散优化问题,同时结合 Lévy 飞行搜索策略,提出了一种新型的红外光谱特征选择算法。采用三个红外光谱数据集对提出的算法进行了验证,同时与遗传算法、模拟退火算法、无信息变量消除法等进行了比较分析。实验结果显示,该方法可以快速地搜索到全局最优值,能有效地提高波长选择的准确性和稳定性,被选择的波长物理、化学意义明确,采用选择的特征波段建立的定量模型优于用全谱建立的模型。同时,三个不同相态、不同光谱范围的数据集表明,所提出的算法具有较大的适用范围与实用价值。

关键词: 特征选择; 蝙蝠算法; Lévy 飞行; 红外光谱

中图分类号: TN219 **文献标志码:** A **文章编号:** 1007-2276(2014)08-2715-07

Feature selection of infrared spectrum based on improved bat algorithm

Chen Yuanyuan^{1,3}, Wang Zhibin^{1,2,3}, Wang Zhaoba^{1,2}

- (1. State Key Laboratory for Electronic Measurement Technology, Taiyuan 030051, China;
2. Key Lab of Instrumentation Science & Dynamic Measurement, Ministry of Education, Taiyuan 030051, China;
3. Engineering Technology Research Center of Shanxi Province for Opto-Electronic Information and Instrument, Taiyuan 030051, China)

Abstract: Feature selection is an important part during the process of qualitative and quantitative analysis of infrared spectrum. In order to solve the disadvantage of traditional methods, such as multi-parameters, slow convergence, poor accuracy, prone to premature, etc., a novel feature selection algorithm was proposed, which combined the basic bat algorithm and Lévy flights search strategy. Meanwhile, due to the original version of bat algorithm was only suitable for continuous problems, a binary version of bat algorithm was proposed. Three infrared spectrum datasets were used to check the performance of proposed method while the comparisons with traditional genetic algorithm, simulate anneal algorithm and uninformative variable elimination methods were also implemented. The experiment results show that, the proposed method can quickly find the global best combination of sub-intervals and improve the accuracy and stability of feature selection. More importantly, the selected wavenumbers have exactly physical meanings. Meanwhile, the

收稿日期: 2013-12-15; 修订日期: 2014-01-10

基金项目: 国家自然科学基金科学仪器基础研究专款(61127015); 国家国际科技合作专项(2012DFA10680、2013DFR10150); 山西省青年科技研究基金(2013021028-1)

作者简介: 陈媛媛(1980-), 女, 讲师, 博士, 主要从事光谱信号处理技术、智能算法方面的研究。Email: chenyy@nuc.edu.cn

generalized performance of the model established based on the selected wavenumbers was better than the whole spectral range. The tests on three different phases (solid, liquid and gas) and different spectral range indicated that, the proposed algorithm has a widely practical scope and value.

Key words: feature selection; bat algorithm; Lévy flights; infrared spectrum

0 引言

随着化学计量学和仪器技术的快速发展, 红外光谱分析作为一种新兴的分析技术, 以其分析速度快、效率高、成本低和易于实现在线分析等特点, 在石油、化工、农业、医药、烟草、食品、环境监测、航空航天等行业取得了广泛应用^[1-3]。由于光谱信息具有波长点多、波长点之间存在严重的共线性关系等特点, 在利用全光谱建立校正模型前, 通常需要进行波长选择操作, 即通过对光谱的适当选择、处理或变换, 剔除无信息变量及噪声, 从而减少校正模型的预测误差和增强模型的可靠性。

目前, 常用的波长选择方法主要有基于光谱的物理意义选择(KBS)、连续投影算法(SPA)、无信息变量消除法(UVE)、模拟退火算法(SA)、遗传算法(GA)、人工神经网络(ANN)、区间偏最小二乘法(iPLS)等, 各自的优缺点如表 1 所示。

表 1 传统特征选择方法优缺点比较

Tab.1 Advantage and disadvantage comparison of traditional variable selection methods

Methods	Advantage	Disadvantage
KBS ^[4]	Suitable for professionals	Need professional knowledge
SPA ^[4-5]	Reduce multicollinearity problem of regression	Low SNR characteristic wavelength
UVE ^[4,6]	To avoid over fitting problem	A large number of remaining variables after UVE
SA ^[4]	To search global optimum	Slow convergence rate
GA ^[4,7]	To ensure diversity of searching space	A large number of adjustable parameters, slow convergence rate
ANN ^[4]	A nonlinear calibration model	"black box" model be hard to explain
iPLS ^[4,8]	Graphical display characteristic interval	A large number of optimized parameters

为了综合上述方法的优点, 有效避免其缺点, 文中提出了一种新的红外光谱特征选择方法。该方法基于蝙蝠算法(BA), 结合 Lévy 飞行搜索策略, 在全局范围内搜寻最佳的波长组合, 为后续建立校正模型提供了有效的特征变量, 提高了建模精度, 缩短了建模时间。

1 算法描述

1.1 改进蝙蝠算法

蝙蝠算法是 Xin-She Yang 于 2010 年提出的一种新型启发式算法。蝙蝠在搜寻猎物过程中, 通过调整发出的超声波脉冲的强度实现大范围的搜索, 通过调整发出的超声波脉冲的频度精确掌握猎物不断变化的空间位置。两者相结合, 可以帮助蝙蝠在黑暗的环境中躲避障碍物且能捕食猎物^[9]。蝙蝠算法是基于群体进化的算法, 首先在可行解空间随机初始化种群, 即确定个体的初始位置和初始速度, 其中位置用于表征问题的可能解; 在搜索过程中, 通过更新个体的飞行速度和位置, 从而精确掌握猎物的空间位置。基本蝙蝠算法的详细步骤可参考文献[9]。

然而, 通过深入分析可以发现, 蝙蝠算法的思想是向当前最优个体学习进行速度更新, 从而实现位置更新。若当前最优个体被局部极值吸引, 并没有有效机制来摆脱束缚, 从而导致群体迅速丧失多样性并失去进化能力, 即基本蝙蝠算法存在易早熟的缺点。

为解决这个问题, 文中引入 Lévy 飞行搜索策略。Lévy 飞行搜索策略模拟了自然界中许多生物觅食时经常采用的短距离的探索性蹦蹦跳跳与偶尔较长距离的行走相间的特点。Lévy 飞行轨迹模拟如图 1 所示。

引入 Lévy 飞行搜索策略后的个体位置更新公式如公式(1)所示^[10]:

$$\mathbf{x}_i^t = \mathbf{x}_i^{t-1} + (\mathbf{x}_i^{t-1} - \mathbf{x}^*) \otimes \mathbf{L}(\lambda) \quad (1)$$

式中: \mathbf{x}_i^{t-1} 、 \mathbf{x}_i^t 分别表示蝙蝠 i 在 $t-1$ 和 t 时刻的空间位置; \mathbf{x}^* 表示在当前群体中最佳蝙蝠所处位置; $\mathbf{L}(\lambda)$

表示跳跃步长服从 Lévy 分布的随机搜索向量, λ ($1 \leq \lambda \leq 3$) 为尺度参数; \otimes 表示矢量运算。

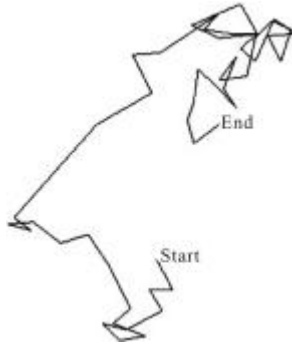


图 1 Lévy 飞行轨迹模拟

Fig.1 Simulation track of Lévy flights

一系列的研究证明:利用 Lévy 飞行搜索策略会产生较大跳跃这种不均匀随机游走的特性,可以有效避免局部极值的吸引。

1.2 基于改进蝙蝠算法的红外光谱特征选择算法

与遗传算法类似,需要将红外光谱的特征选择映射为一个目标函数的优化问题。文中进行如下处理:

首先,选择校正模型的验证集的均方根误差 (RMSE) 作为目标函数,由于传统的 PLS 在建立非线性关系映射时具有一定的局限性,因此,文中采用由 Huang 等提出的极限学习机^[11] (ELM) 实现,ELM 本质上是一种三层的前馈神经网络,因此,处理非线性映射具有一定的优势,但与传统 BP 神经网络不同,ELM 的训练过程无需迭代,仅需计算一个线性方程组即可,另外,ELM 具有可调参数少的特点,只有隐含层神经元个数这一个参数需要设置。

其次,将整个光谱范围划分为若干个子区间,蝙蝠个体用一串二进制码表示,对应子区间的选择与否则用“1”与“0”表示。由于基本的蝙蝠算法针对的是连续优化问题,因此文中利用公式(2)和(3)对个体的位置进行离散化处理^[12]。

$$V(v_i^t(k)) = \frac{2}{\pi} \arctan\left(\frac{2}{\pi} v_i^t(k)\right) \quad (2)$$

$$x_i^t(k+1) = \begin{cases} (x_i^t(k+1))^{-1}, & \text{if } \text{rand} < V(v_i^t(k)) \\ x_i^t(k+1), & \text{if } \text{rand} \geq V(v_i^t(k)) \end{cases} \quad (3)$$

式中: $v_i^t(k)$ 表示蝙蝠 i 在 t 时刻第 k 维方向上的速度; $x_i^t(k)$ 表示蝙蝠 i 在 t 时刻第 k 维方向上的位置; $(x_i^t(k+1))^{-1}$ 是 $x_i^t(k)$ 的二进制反码; rand 为 $[0, 1]$ 范围

内的随机数。

综上所述,基于改进蝙蝠算法(LBA)的红外光谱特征选择算法流程如下。

(1) 初始化:设置基本参数,随机初始化种群(蝙蝠数目 m) 的位置 x_i 与速度 v_i , 针对每个蝙蝠,利用 ELM 建立校正模型,计算验证集的 RMSE 并找出群体中处于最佳位置 x^* 的个体;

(2) 更新个体位置:按照公式(1)更新蝙蝠当前位置,并根据公式(2)和(3)进行离散化处理;

(3) 局部搜索:生成随机数 R_1 , 若 $R_1 > r_1$, 从当前种群中选择最佳解,并在该解的附近随机产生一个局部解;同时,生成随机数 R_2 , 若 $R_2 < A_1$ 并且蝙蝠当前位置得到改善,则飞至更新后的位置,并调整脉冲频率 r_1 和脉冲强度 A_1 ;

(4) 根据校正模型的验证集的 RMSE 对蝙蝠群体进行评估,找出当前最佳蝙蝠及所处空间位置;

(5) 迭代次数达到最大迭代次数 N 转入(6), 否则转入(2), 进行下一次搜索;

(6) 输出最佳的光谱子区间组合。

2 实验

为了验证方法的有效性和优越性,采用了 3 个不同相态、不同光谱范围的数据集进行了实验。所有数据集的样本均分成训练集、测试集和验证集,其中训练集用于建立校正模型,验证集用于计算 RMSE 并进行特征选择,测试集用于评价方法的优劣。

2.1 数据集

3 个数据集情况如下:

(1) 玉米的近红外光谱与植物油含量。光谱采集范围是 1 100~2 498 nm, 步长 2 nm。训练集 50 个样本,测试集和验证集各 15 个样本。该数据集可以从 Eigenvector Research 公司的网站下载^[1]。

(2) 汽油的近红外光谱和辛烷值。光谱采集范围是 900~1 700 nm, 步长 2 nm。为了便于与参考文献[1] 进行比较,将前 100 个几乎不包含信息的波长(即 900~1 100 nm) 剔除。训练集 40 个样本,测试集和验证集各 10 个样本。该数据集可以从 MATLAB 软件自带的数集中获取(spectra.mat)。

(3) 混合气体的红外光谱和 CO、N₂O 的浓度。选用加拿大 ABB MR170 FTIR 红外光谱仪进行实验数据采集,外接气室长度 10 cm。环境条件为温度

296 K, 气压 1 013.25 hPa, 将不同浓度的 CO、CO₂、N₂O 和 N₂ 气体注入气室, 浓度控制采用不同精度流量计来实现。光谱采集范围是 2 140~2 220 cm⁻¹, 步长 0.5 cm⁻¹。训练集 240 个样本, 测试集和验证集各 80 个样本。

需要说明的是, 玉米与汽油数据集虽然样本规模较小, 但均是光谱分析中常用的“标准”光谱, 样本的代表性较好。混合气体数据集由严格的实验操作采集而得, 经过仔细检查, 异常样本被排除出该数据集。因此, 可以认为文中选择的 3 个数据集中均不存在异常样本, 不会给特征光谱的筛选及校正模型的建立带来误差与影响。

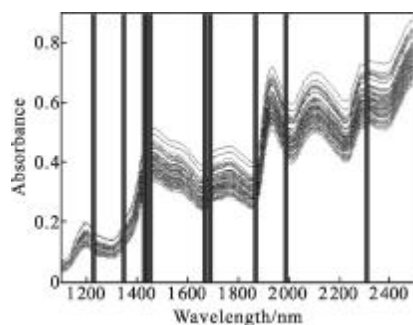
2.2 实验结果

各数据集实验结果如下。

(1) 玉米数据集

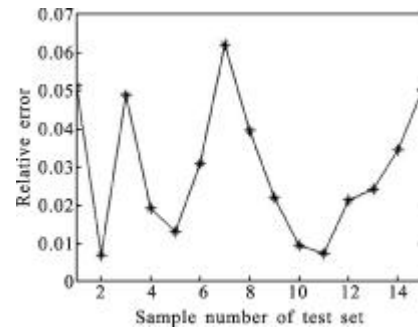
玉米数据集的全光谱范围内共包含 700 个波长点, 文中将整个光谱范围均匀划分成 70 个子区间, 每个子区间包含 10 个波长点。算法的基本初始参数设置如下: 蝙蝠数目 $m=30$, 最大迭代次数 $N=50$, 脉冲频率范围 $[f_{\min}, f_{\max}]=[0, 2]$, 脉冲强度 $A=0.9$, 脉冲频率 $r=0.9$, Lévy 分布尺度参数 $\lambda \approx 1.5$ 。

利用 LBA 优化搜索出的特征波长子区间组合如图 2(a)所示, 共包含 9 个子区间共 90 个波长点。参考文献[1]中的 MC-VS、GA、UVE-PLS 三种方法选择的波长点数分别是 144、134 和 291。对比可以发现, LBA 的搜索结果覆盖了主要的吸收波段, 如 1 200 nm 附近 C-H 二级倍频吸收区, 1 600~1 800 nm C-H 伸缩振动一级倍频吸收区, 2 000 nm、2 300 nm 附近 C-H 合频吸收区, 1 430 nm 附近 O-H 伸缩振动一级倍频吸收区等, 同时进一步减少了波长点个数, 利用该结果建立了植物油的 ELM 校正模型如



(a) 玉米数据集的优化波长子区间

(a) Corn dataset's selected wavenumber sub-intervals



(b) 玉米数据集的测试集预测误差

(b) Corn dataset's predictive error of test set

图 2 玉米数据集结果

Fig.2 Corn dataset's results

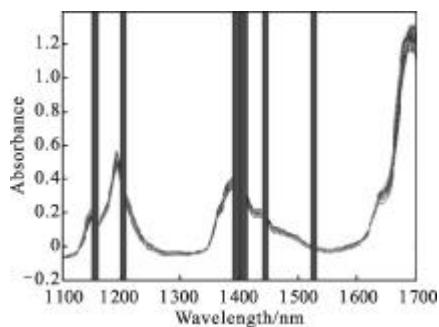
图 2(b)所示, 经测试集验证, RMSEP 为 0.121 6, 与参考文献[1]中 MC-VS 和 GA 预测性能相当, 表明经过 LBA 方法的优化筛选即保证模型预测精度的同时降低了模型的复杂度。

(2) 汽油数据集

汽油数据集的全光谱范围内共包含 300 个波长点, 整个光谱范围被均匀划分成 50 个子区间, 每个子区间包含 6 个波长点。算法的基本初始参数设置如下: 蝙蝠数目 $m=20$, 最大迭代次数 $N=30$, 脉冲频率范围 $[f_{\min}, f_{\max}]=[0, 2]$, 脉冲强度 $A=0.9$, 脉冲频率 $r=0.9$, Lévy 分布尺度参数 $\lambda \approx 1.5$ 。

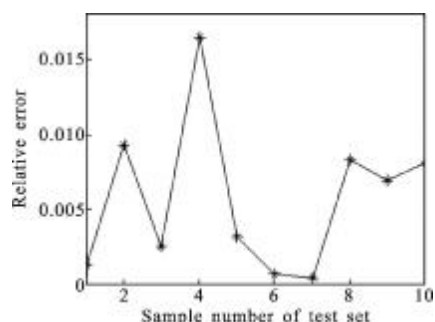
利用 LBA 优化搜索出的特征波长子区间组合如图 3(a)所示, 共包含 6 个子区间共 36 个波长点, 所选特征波长涵盖了 1 151 nm 附近芳烃和 CH₃ 伸缩振动二级倍频吸收区, 1 194 nm 附近 CH₃ 伸缩振动二级倍频吸收区, 1 394 nm 附近 CH₂ 合频吸收区, 1 412 nm 附近芳烃和 CH₂ 的合频吸收区, 1 435 nm 附近芳烃和 CH 的合频吸收区等主要吸收波段。与参考文献[1]中的三种方法相比可以发现: MC-VS 选择的波长点数最少, 仅有 12 个, 但结果没有涵盖主要吸收波段; LBA 次之, 包含 36 个波长点, 在包含主要吸收波段情况下点数最少; 而 GA 和 UVE-PLS 则分别包含 68 和 77 个波长点, 筛选效果不明显。

利用优化的波长子区间组合建立的校正模型的测试集相对预测误差如图 3(b)所示, 对应的 RMSEP 为 0.140 6, 均小于参考文献[1]中的三种方法(分别为 0.169 5, 0.167 1 和 0.157 9)。综合考虑选择的波长点数与校正模型的预测性能, 可以认为 LBA 方法有效地筛选出了特征波长。



(a) 汽油数据集的优化波长子区间

(a) Gasoline dataset's selected wavenumber sub-intervals



(b) 汽油数据集的测试集预测误差

(b) Gasoline dataset's predictive error of test set

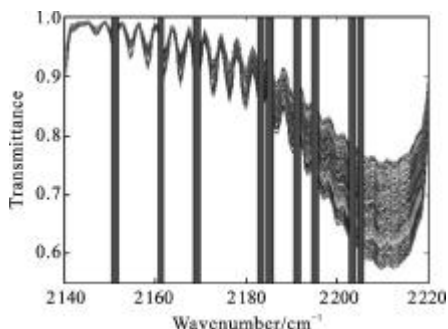
图 3 汽油数据集结果

Fig.3 Gasoline dataset's results

(3) 混合气体数据集

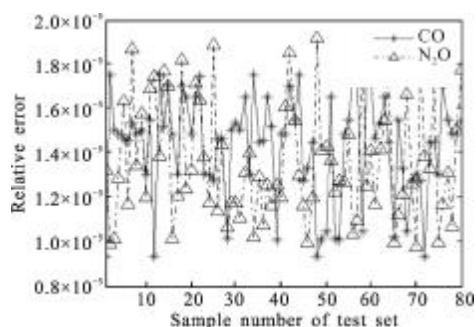
整个光谱范围内共 160 个波长点,被均匀地划分成 40 个子区间,每个子区间包含 4 个波长点。算法的基本初始参数设置如下:蝙蝠数目 $m=10$,最大迭代次数 $N=20$,脉冲频率范围 $[f_{min}, f_{max}]=[0, 2]$,脉冲强度 $A=0.9$,脉冲频度 $r=0.9$,Lévy 分布尺度参数 $\lambda=1.5$ 。

图 4(a)给出了 LBA 优化搜索到的最佳波长子区间组合,共包含 9 个子区间,36 个波长点。CO 气体的



(a) 混合气体数据集的优化波长子区间

(a) Mixed gases dataset's selected wavenumber sub-intervals



(b) 混合气体数据集的测试集预测误差

(b) Mixed gases dataset's predictive error of test set

图 4 混合气体数据集结果

Fig.4 Mixed gases dataset's results

强吸收带为 $2160\sim 2180\text{ cm}^{-1}$, N_2O 气体的强吸收带为 $2180\sim 2220\text{ cm}^{-1}$, LBA 优化搜索结果与之一致。

利用 LBA 筛选出的特征波长点建立的校正模型测试集相对预测误差如图 4(b)所示,CO 与 N_2O 气体的测试集 RMSEP 分别为 0.176 8 和 0.193 2, 呈现出了较好的泛化性能与预测精度。

2.3 分析与讨论

2.3.1 算法收敛速度分析

由于参考文献[1]没有给出仿真平台的详细配置,文中提出的 LBA 方法无法进行严谨的对比。笔者自行编写了 MC-VS,GA 与 SA 的 MATLAB 程序,并在 Windows64 位操作系统、CPU3.00 GHz、4.00 GB RAM 计算机上与 LBA 进行对比。以玉米数据集为例,某次运行对应的迭代进化曲线如图 5 所示。

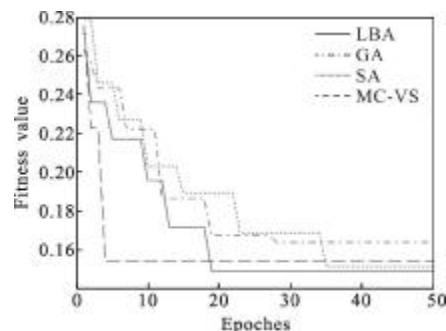


图 5 迭代进化过程对比

Fig.5 Comparison of iterative evolution process

从图中可以直观地看出,MC-VS 速度最快,仅需 4 次迭代(0.3 s)便可收敛,LBA 速度次之,需要 19 次迭代(1.2 s)达到收敛,GA 需迭代 28 次(3.4 s)才能收敛,SA 收敛速度最慢,需要迭代 35 次(5.1 s)。这是

因为,与启发式搜索算法相比,MC-VS 算法利用蒙特卡洛随机模拟,具有一定的速度优势;遗传算法种群规模较大导致收敛速度较慢,而模拟退火算法的冷却过程太长。同时,通过对比最终寻找到的最佳适应度函数值可以发现,LBA 要显著优于 MC-VS,GA 和 SA,可以得出结论,LBA 是一个有效的搜索方法。

2.3.2 算法寻优稳定性分析

与遗传算法类似,蝙蝠算法具有一定的随机性,即每次运行的结果都不相同,给算法的稳定性评价造成了一定的困难。为了对算法的寻优稳定性进行有效的分析,这里提出了一个新的评价指标--特征选择因子,其定义如公式(4)所示:

$$SF = \frac{N_1 + N_2 + N_3}{N} \quad (4)$$

$$N_1 = \{x \in S_1 \cap x \in S_2\}$$

$$N_2 = \{x \in S_1 \cap x \in S_2 + \delta\}$$

$$N_3 = \{x \in S_2 \cap x \in S_1 + \delta\}$$

式中: N 为全部波长点的个数; N_1 为在两次运行结果中均被选择的波长点个数; N_2 为在第一次运行结果中被选择,第二次运行结果中对应位置邻域内亦存在被选择的波长点个数; N_3 为在第二次运行结果中被选择,第一次运行结果中对应位置邻域内亦存在被选择的波长点个数。 SF 的值介于 0 和 1 之间,越趋近于 1 表示算法的寻优稳定性越好,反之,表明算法的寻优稳定性越差。

表 2 列出了 3 个数据集的 LBA 与参考文献 [1] 中的三种算法在运行 100 次情况下,两两计算 SF 后取平均值的结果,邻域范围的宽度设置为 2,即若在第一次运行结果中第 i 个特征被选择,在第二次运行结果中第 $i-1$ 或 $i+1$ 个特征也被选择,则计数加 1。从表 2 中可以看出,LBA 的寻优稳定性要明显优于 MC-VS、GA、SA 和 UVE-PLS,即表明 LBA 的可重复性更佳。

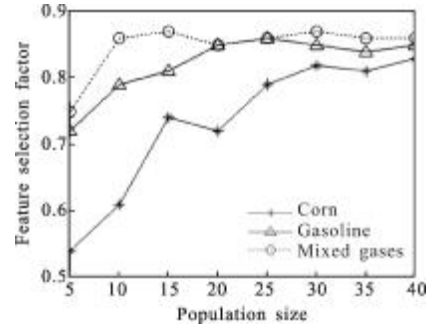
表 2 LBA 与 GA 等算法的特征选择因子比较

Tab.2 Feature selection factor comparison of several algorithms

Algorithm	Corn	Gasoline	Mixed gase
LBA	0.82	0.85	0.87
MC-VS	0.75	0.81	0.83
GA	0.76	0.75	0.73
SA	0.78	0.79	0.75
UVE-PLS	0.72	0.70	0.76

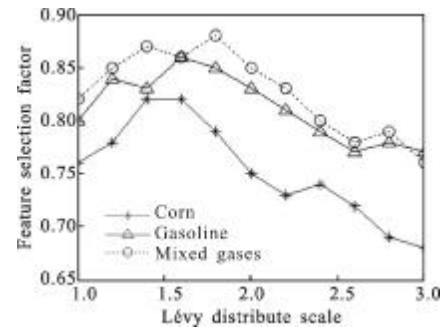
2.3.3 算法参数设置分析

LBA 算法主要参数为:种群规模大小及 Lévy 分布尺度参数,这两个参数对算法寻优稳定性的影响分别如图 6(a)与图 6(b)所示。



(a) 种群规模对特征选择因子的影响

(a) Influence of population size on feature selection factor



(b) Lévy 分布尺度参数对特征选择因子的影响

(b) Influence of Lévy distribute scale on feature selection factor

图 6 算法参数对寻优稳定性的影响

Fig.6 Influence of parameter settings on optimization stability

图 6(a)为 3 个数据集中种群规模大小对性能稳定性的影响,从图中可以看出,当种群规模较小时,即在整个解空间中执行搜索任务的蝙蝠个体较少,寻优稳定性较差,而当种群规模达到一定数量时,蝙蝠个体的增加对寻优稳定性几乎没有影响,但是计算量将会增大。一般情况下,遗传算法为了保证种群的多样性,种群的规模通常设置较大(大于 100)。然而,蝙蝠算法结合 Lévy 飞行搜索策略,则无需设置太多蝙蝠个体(通常设 10~30 即可),即可以实现整个解空间的全局搜索。但是,具体的参数取值还需结合问题本身进行确定,例如,从图 3(a)中可以清晰地看到,位于 1520 nm 附近的子区间并非特征峰的位置,然而 LBA 却将其错误地选入。从图 6(a)中可以看出,种群规模达到 20 以上时,寻优结果已趋于稳定,当笔者将蝙蝠数目从 20 增加到 30 后,该问题即

得到了有效的避免。

图 6(b)为 3 个数据集中 Lévy 分布尺度参数对性能稳定性的影响,从图中可以看出,当 Lévy 分布尺度参数取值 1.5 左右时,寻优稳定性最好。一方面,随着 Lévy 分布尺度参数的逐渐增加,蝙蝠跳跃的步长逐渐减小,因此 Lévy 飞行搜索策略的优势将逐渐消失,此时演变成为局部搜索,容易陷入局部极小值。另一方面,随着 Lévy 分布尺度参数的逐渐减小,蝙蝠跳跃的步长逐渐增大,此时若种群规模较小,则搜索过程将演变成少许的蝙蝠在整个解空间中作长距离跳跃,基本蝙蝠算法的局部搜索的能力将被减弱。

因此,在解决特定的问题时,应该综合平衡考虑种群规模及 Lévy 分布尺度参数的取值。

3 结 论

文中在基本蝙蝠算法的基础上,充分利用 Lévy 飞行搜索策略在整个解空间中随机游走的特性,提出了一种新的红外光谱特征选择方法。3 个不同相态、不同光谱范围的数据集验证结果表明,该方法可以寻找到全局最佳特征波长组合,与传统算法相比,具有收敛速度快、全局寻优等优点,具有广泛的应用范围与实用价值。

参考文献:

- [1] Hong Mingjian, Wen Quan, Wen Zhiyu. New near infrared wavelength selection algorithm based on Monte-Carlo method [J]. *Acta Optica Sinica*, 2010, 30(12): 3637-3642. (in Chinese)
洪明坚, 温泉, 温志渝. 一种基于蒙特卡罗方法的近红外波长选择算法[J]. *光学学报*, 2010, 30(12): 3637-3642.
- [2] Qin Yuhua, Ding Xiangqian, Gong Huili. High dimensional feature selection in near infrared spectroscopy classification [J]. *Infrared and Laser Engineering*, 2013, 42(5): 1355-1359. (in Chinese)
秦玉华, 丁香乾, 宫会丽. 高维特征选择方法在近红外光谱分类中的应用[J]. *红外与激光工程*, 2013, 42(5): 1355-1359.
- [3] Guo Tengxiao, Ding Xuequan, Dong Xiaoqiang, et al. New method of preprocessing IR remote sensing spectrum signals based on EMD [J]. *Infrared and Laser Engineering*, 2013, 42(12): 3196-3200. (in Chinese)
郭腾霄, 丁学全, 董晓强, 等. 基于 EMD 的红外遥测光谱信号预处理新方法 [J]. *红外与激光工程*, 2013, 42(12): 3196-3200.
- [4] Zou Xiaobo, Zhao Jiwen, Povey M J W, et al. Variables selection methods in near-infrared spectroscopy[J]. *Analytica Chimica Acta*, 2010, 667: 14-32.
- [5] Liu Fei, Jiang Yihong, He Yong. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: a case study to determine soluble solids content of beer [J]. *Analytica Chimica Acta*, 2009, 635(1): 45-52.
- [6] Cai Wensheng, Li Yankun, Shao Xueguang. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra [J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 90(2): 188-194.
- [7] Givianrad M H, Saber-Tehrani M, Zarin S. Genetic algorithm-based wavelength selection in multicomponent spectrophotometric determinations by partial least square regression: application to a sulfamethoxazole and trimethoprim mixture in bovine milk [J]. *Journal of The Serbian Chemical Society*, 2013, 78(4): 555-564.
- [8] Cramer J A, Kramer K E, Johnson K J, et al. Automated wavelength selection for spectroscopic fuel models by symmetrically contracting repeated unmovable window partial least squares [J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, 92: 13-21.
- [9] Yang Xinshe. A new metaheuristic bat-inspired algorithm [J]. *Nature Inspired Cooperative Strategies for Optimization*, 2010, 284: 65-74.
- [10] Liu Changping, Ye Chungming. Bat algorithm with the characteristics of Lévy flights [J]. *CAAI Transactions on Intelligent Systems*, 2013, 8(3): 240-246. (in Chinese)
刘长平, 叶春明. 具有 Lévy 飞行特征的蝙蝠算法[J]. *智能系统学报*, 2013, 8(3): 240-246.
- [11] Huang Guangbin, Zhu Qinyu, Siew C K. Extreme learning machine: theory and applications [J]. *Neurocomputing*, 2006, 70(1): 489-501.
- [12] Mirjalili S, Mirjalili S M, Yang Xinshe. Binary bat algorithm [J]. *Neural Computing and Applications*, 2013: 1-19.