

思维进化蝙蝠算法及其在混合气体红外光谱特征选择中的应用

陈媛媛^{1,2}, 王志斌^{1,2,3}, 王召巴^{1,3}

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

摘要: 为了解决混合气体多组分间特征吸收峰相互重叠引起的特征选择困难问题, 提出了新型红外光谱特征选择方法, 并对该方法的性能进行了分析与评价。首先, 充分结合思维进化计算的并行机制、异化操作与蝙蝠算法的局部搜索能力, 设计了思维进化蝙蝠算法。接着, 通过实验采集两个混合气体数据库, 利用思维进化蝙蝠算法对其目标组分的特征峰进行筛选。然后, 从算法的收敛速度和筛选出的特征峰两个方面, 将思维进化蝙蝠算法与基本蝙蝠算法、遗传算法、粒子群优化算法及并行萤火虫群优化算法等进行比较。最后, 讨论了思维进化蝙蝠算法与无信息变量消除法相结合对结果的影响。实验结果表明: CO的特征峰范围包括 $2\ 090\sim 2\ 110\ \text{cm}^{-1}$ 和 $2\ 115\sim 2\ 125\ \text{cm}^{-1}$, 共包含 32 个波长点; N_2O 的特征峰范围为 $2\ 225\sim 2\ 250\ \text{cm}^{-1}$, 共包含 26 个波长点。利用筛选出的特征波长点建立的浓度反演模型, 测试集均方根误差为 0.155, 决定系数可达 0.908。实验结果表明: 思维进化蝙蝠算法收敛速度快、全局搜索能力强, 适用于存在重叠特征峰的混合气体的特征选择, 对应的浓度反演模型的泛化性能也有显著提升。

关键词: 特征选择; 思维进化计算; 蝙蝠算法; 混合气体; 红外光谱

中图分类号: TN219 **文献标志码:** A **文章编号:** 1007-2276(2015)03-0845-07

Mind evolutionary bat algorithm and its application to feature selection of mixed gases infrared spectrum

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

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

Abstract: Due to the fact that the characteristic peaks of multi-component of mixed gases have overlapping problem, it was hard to implement feature selection for each target gas. To solve this problem, a novel feature selection method was introduced. First, by making full use of the parallel mechanism, dissimilation operator of mind evolutionary computation and local search ability of bat algorithm, the mind evolutionary bat algorithm was designed. Two different mixed gases databases were

收稿日期: 2014-07-11; 修订日期: 2014-08-12

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, 副教授, 博士, 研究方向为光谱信号处理技术、智能算法。Email: chenyy@nuc.edu.cn

collected to validate the performance of proposed method. Then, from the aspects of convergence speed and characteristic peaks, the comparison with basic bat algorithm, genetic algorithm, particle swarm optimization and parallel glowworm swarm optimization algorithm was investigated. Finally, the influence of combination with uninformative variable elimination method was discussed. Experimental results show that the characteristic peaks of carbon monoxide include $2\ 090\text{--}2\ 110\ \text{cm}^{-1}$ and $2\ 115\text{--}2\ 125\ \text{cm}^{-1}$, which total have 32 wavelength points while the characteristic peaks of nitrogen oxide were in range from $2\ 225$ to $2\ 250\ \text{cm}^{-1}$, which total have 26 wavelength points. Considering the concentration retrieve model established with the selected characteristic peaks, the root mean squared error of prediction set was 0.155, and the determined coefficient can reach as high as 0.908. Experimental results show that the proposed method has the advantage of rapid convergence speed and well global search ability, which was adaptable to do the feature selection for those mixed gases with overlapping problem.

Key words: feature selection; mind evolutionary computation; bat algorithm; mixed gases; infrared spectrum

0 引言

近年来,随着化学计量学和仪器技术的快速发展,红外光谱分析以其分析速度快、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]。其中,在环境检测领域,红外光谱分析技术已经在各种典型的大气科学应用场合,如:精确测定环境大气中的多种痕量组分、地表与大气痕量气体交换的微气象测量、生物燃烧气体、气体示踪法测量痕量气体的局地排放等方面,发挥出一定的作用,显示了广阔的应用前景。但是,在对大气中某一痕量气体的浓度进行检测时,经常遇到以下困难:一方面,往往大气中存在的其他气体在其分析的波段也有吸收,因此直接测量得到的光谱大多是混合光谱,给后续浓度反演模型的建立造成影响;另一方面,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。另外,在实际遥测环境中,由于光程较长,经常出现气体强吸收峰饱和的情况,无法直接利用气体的强吸收峰建立浓度反演模型,需要借助一些弱吸收峰,但弱吸收峰经常与其他气体组分的吸收峰相互重叠,给后续的分析造成很大的困难。因此,在建立浓度反演模型前,通常需要对光谱进行选择,剔除无信息波长点及噪声,从而减少模型的预测误差和增强模型的可靠性^[4]。然而,传统的一些特征选择算法,均具有一定的局限性。例如,区间偏最小二乘法(iPLS)依赖于子区间的窗口宽度且是

单向叠加式选择;无信息变量消除法(UVE)筛选后的波长点数仍然较多;连续投影算法(SPA)选择的波长点信噪比低;模拟退火算法(SA),遗传算法(GA)与粒子群优化(PSO)算法收敛速度慢且容易陷入局部极小值^[5]。

上述局限性从一定程度上讲,是由两个因素造成的。其一,光谱波长点数较多引起的高维优化问题;其二,混合气体光谱中各个气体的特征峰存在相互重叠的问题。因此,为了解决上述问题,文中结合思维进化算法与蝙蝠算法,提出了一个高效、并行的特征选择算法——思维进化蝙蝠算法。该算法充分利用思维进化计算的并行机制来实现快速搜索;利用思维进化计算的异化操作保证种群的多样性,即扩大整个解空间的搜索范围,提升全局搜索能力,减小陷入局部极小的概率;利用蝙蝠算法的局部搜索能力,来获得局部范围的最优解。

1 算法描述

1.1 基本蝙蝠算法

蝙蝠算法是由杨新社于 2010 年提出的一种生物启发式算法^[6],该算法基于微型蝙蝠的回声定位特点,通过模拟蝙蝠在觅食过程中脉冲发射频率与脉冲响度的变化规律,利用一个可调节频率技术实现解空间搜索范围的扩大,同时利用自动缩放实现搜索过程中全局搜索与局部搜索间的平衡。

假设某个蝙蝠在 t 时刻的速度为 v_t ,位置为 x_t ,当前种群所有蝙蝠中,最佳的解为 x^* 。根据回声定位

的思想,蝙蝠的速度和位置更新公式如下:

$$f_i = f_{\min} + (f_{\max} - f_{\min})\beta \quad (1)$$

$$v_i' = v_i^{t-1} + (x_i^{t-1} - x_i^*)f_i \quad (2)$$

$$x_i' = x_i^{t-1} + v_i' \quad (3)$$

式中: f_{\min} 和 f_{\max} 分别为脉冲发射最小和最大频率; $\beta \in [0, 1]$ 是一个服从均匀分布的随机向量。

为了提供一个能够控制全局搜索与局部搜索的有效控制,蝙蝠算法在迭代搜索过程中引入自适应的脉冲发射响度 A_i 与脉冲发射速率 r_i 。通常,当蝙蝠寻找到食物时,脉冲发射响度减小,脉冲发射速率增加。因此,脉冲发射响度可以在最小响度 A_{\min} 和最大响度 A_{\max} 间任意取值。假设 $A_{\min}=0$ 表示蝙蝠刚寻找到食物并停止发射脉冲。由以上假设可得:

$$A_i^{t+1} = \alpha A_i^t \quad (4)$$

$$r_i^{t+1} = r_i^0 [1 - \exp(-\gamma t)] \quad (5)$$

其中, α 与 γ 是常数。从意义上看, α 与模拟退火算法中的冷却因子类似。对于任意的 $0 < \alpha < 1$ 和 $\gamma < 0$,可以推出:

$$A_i^t \rightarrow 0, r_i^t \rightarrow r_i^0, \text{ as } t \rightarrow \infty \quad (6)$$

1.2 思维进化计算

思维进化计算最初由孙承意等提出,其基本架构如图 1 所描述。

与遗传算法不同的地方在于:首先,思维进化计算中的种群被划分成优胜子群体与临时子群体;其次,思维进化计算有一个全局公告板,每个子群体包含若干个局部公告板。全局公告板用于实现优胜子群体与临时子群体的竞争,局部公告板用于实现子群体的局部搜索。受人类思维方式的启发,思维进化计算设计了两种操作算子:趋同与异化。趋同算子在

局部解空间中执行局部搜索,异化算子在整个解空间中执行全局搜索。思维进化计算的详细理论及算法流程见参考文献[7]。

1.3 思维进化蝙蝠算法

在利用该算法进行特征选择前,首先需要将光谱的特征波长选择问题映射为优化问题。文中进行如下处理:每一个蝙蝠个体(可能解)用一串与整个光谱范围波长点数相等的二进制码表示,各个波长点的选择与否用“1”与“0”表示。由于基本的 BA 算法针对的是连续优化问题,因此无法直接应用到“0-1”优化问题中。利用 v 型传递函数^[8]进行离散化处理,对应的位置更新公式如下:

$$V(v_i^k(t)) = \frac{2}{\pi} \arctan\left(\frac{2}{\pi} v_i^k(t)\right) \quad (7)$$

$$x_i^k(t+1) = \begin{cases} (x_i^k(t))^{-1}, & \text{rand} < V(v_i^k(t+1)) \\ x_i^k(t), & \text{rand} \geq V(v_i^k(t+1)) \end{cases} \quad (8)$$

式中: $x_i^k(t)$ 和 $v_i^k(t)$ 分别为第 i 个蝙蝠在 t 轮迭代时的位置和速度; k 是蝙蝠的维度; $(x_i^k(t))^{-1}$ 为 $x_i^k(t)$ 的补数。

其次,需要确定评价个体性能优劣的指标。文中选择校正模型的测试集均方根误差(RMSEP)作为目标函数。同时,为了克服 PLS 在建立非线性关系映射时的局限性,文中采用由 Huang 等提出的极限学习机^[9](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。

综上所述,基于思维进化蝙蝠算法(MEBA)的红外光谱特征选择算法的详细流程如下:(1)初始化。随机产生初始种群,并划分为若干优胜子群体和临时子群体。(2)趋同操作与局部搜索。在每个子群体内部,通过计算 RMSEP,各个蝙蝠个体相互竞争以产生局部最优解。在初次迭代时,所有个体均散落在优胜个体的周围并服从正态分布。若某个个体的得分高于最优个体的得分,则该个体被视为新的最优个体,同时在局部公告板更新该个体的信息。上述过程重复进行直至子群体成熟。(3)异化操作与全局竞争。若某个临时子群体的得分高于某个优胜子群体的得分,该临时子群体替换优胜子群体。否则,临时子群体将被解散,同时在解空间中随机产生一个新

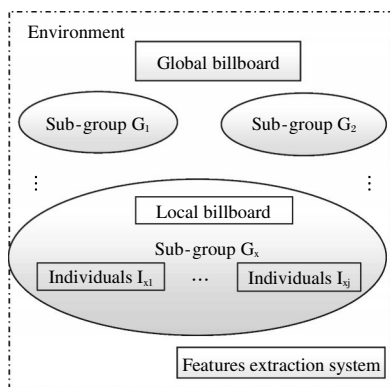


图 1 思维进化计算的架构

Fig.1 Architecture of MEC

的临时子群体。(4) 如果当前迭代次数大于最大迭代次数, 转到 Step5; 否则, 转到 Step2, 开始新一轮迭代搜索。(5) 输出最优个体: 最佳的特征波长点组合。

2 实验方案

为了验证 MEBA 特征选择方法的性能, 文中利用两个混合气体红外光谱数据库(分别用 Mixture #1 数据库和 Mixture #2 数据库表示) 进行测试。其中, Mixture #1 数据库为 CO 和 CO₂ 的混合物, Mixture #2 数据库为 CO、N₂O 和 CO₂ 的混合物。在两个数据库中, CO₂ 均被视为干扰气体。值得强调的是, 在 Mixture #2 数据库中, 文中将待分析的光谱范围限制为 2 150~2 250 cm⁻¹, 该区间内不包含 CO 位于 2 123 cm⁻¹ 附近的强吸收峰, 以此来模拟 CO 的强吸收峰饱和, 弱吸收峰与 N₂O 相互重叠的情况。

每个数据库均被划分成训练集与测试集, 其中训练集用于建立定量浓度反演模型, 测试集用于计算目标函数 RMSEP。Mixture #1 数据库共包含 160 个样本, 其中 120 个样本被选为训练集, 剩余 40 个作为测试集; Mixture #2 数据库共包含 150 个样本, 其中 100 个样本被选为训练集, 剩余 50 个作为测试集。

在此研究中, 两个数据库的数据均是利用由北京瑞利分析仪器有限公司生产的 WQF-520 系列 FTIR 光谱仪采集而得。该光谱仪包含一个角镜型迈克尔逊干涉仪和一个内置的测量范围为 500~5 000 cm⁻¹ 的 DTGS 探测器。综合分析 CO、CO₂ 和 N₂O 的特征吸收峰, 文中将光谱采集范围设置为从 2 000~2 500 cm⁻¹。另外, 折中考虑分辨率与信噪比, 这里将分辨率设置为 1 cm⁻¹。实验在标准环境(1 个大气压, 296 K)下进行, 密闭气室的长度为 10 cm, 针对每个样本, 测量 32 次取平均值作为最终的采集光谱。

3 结果与讨论

3.1 实验结果

3.1.1 Mixture #1 数据库

该组实验中, MEBA 的初始参数设置如下: 整个种群包含 50 只蝙蝠, 被划分成 5 个优胜子群体和 5 个临时子群体, 每个子群体包含 5 只蝙蝠; 最大迭代次数为 20; 脉冲发射频率范围为 [0, 2]; 脉冲响度与发射速率均为 0.9。

MEBA 与 GA 筛选出的特征波长点组合如图 2 所示。从图中可以看出, MEBA 筛选出的 CO 特征峰包含 88 个波长点, 对应的波数范围包括 2 085~2 100 cm⁻¹, 2 115~2 125 cm⁻¹ 及 2 160~2 220 cm⁻¹, 而 GA 筛选出的特征波长点较多, 包含了 2 226~2 237 cm⁻¹ 的冗余信息。如前文所述, 在实验中, CO₂ 被视作是干扰组分, 因此 MEBA 与 GA 均剔除了 2 300~2 400 cm⁻¹ 范围内 CO₂ 的特征峰。上述结果表明: 针对不存在特征峰重叠的混合气体, MEBA 与 GA 均可以有效地剔除干扰波长点, 筛选出目标组分的特征波长点组合, 但 MEBA 筛选出的特征波长点个数更少, 包含的冗余信息更少。

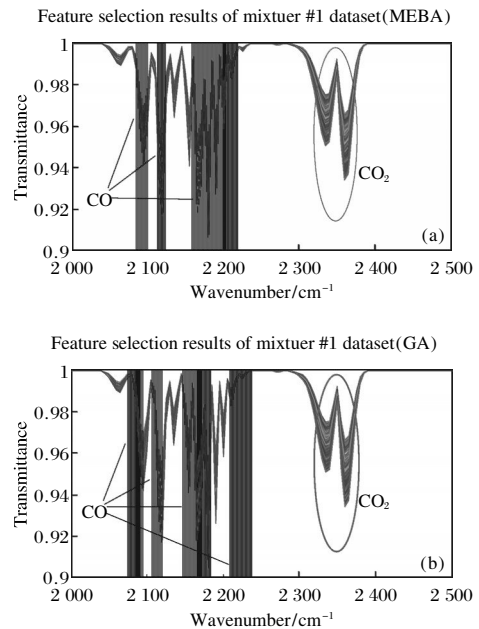


图 2 Mixture #1 数据库的特征选择结果

Fig. 2 Feature selection results of Mixture #1 dataset

3.1.2 Mixture #2 数据库

该组实验中, MEBA 的初始参数设置如下: 整个种群包含 100 只蝙蝠, 被划分成 5 个优胜子群体和 5 个临时子群体, 每个子群体包含 10 只蝙蝠; 最大迭代次数为 20; 脉冲发射频率范围为 [0, 2]; 脉冲响度与发射速率均为 0.9。

与 Mixture #1 数据库不同, Mixture #2 数据库中的目标组分 CO 与 N₂O 存在特征峰重叠的现象。由于 Mixture #1 数据库的实验结果已经表明, 在不存在特征峰重叠的情况下, MEBA 可以有效剔除干扰光谱, 因此这里不再考虑 CO₂ 的干扰, 仅讨论 CO 与 N₂O 的特征峰重叠情况, 对应地将光谱分析范围缩

小至 2 150~2 250 cm⁻¹。

MEBA 与 GA 筛选出的特征波长点组合如图 3 所示,从中可以发现:

(1) MEBA 筛选出的 CO 特征峰包含 12 个波长点,对应的波数范围包括 2 170~2 176 cm⁻¹ 和 2 180~2 184 cm⁻¹,该范围包含 2 173 cm⁻¹ 附近的 CO 弱吸收带。

(2) MEBA 筛选出的 N₂O 特征峰包含 13 个波长点,对应的波数范围为 2 230~2 242 cm⁻¹,该范围覆盖了 N₂O 在 2 237 cm⁻¹ 附近的强吸收带。

(3) 波数 2 190~2 225 cm⁻¹ 是 CO 与 N₂O 的特征峰重叠区域。2 224 cm⁻¹ 附近的吸收峰作为 N₂O 的强吸收带在该区间内,受相互重叠的影响,MEBA 与 GA 算法均没有将其选择。但从图 3 中可以看出,2 205~2 215 cm⁻¹ 范围内的光谱,当采用 GA 算法进行筛选时被选择,而当采用 MEBA 算法进行筛选时未被选择。这表明,针对存在吸收峰重叠的混合气体,MEBA 可以有效地避免重叠区域特征峰的影响,从而筛选出目标组分的独立的特征波长点组合。

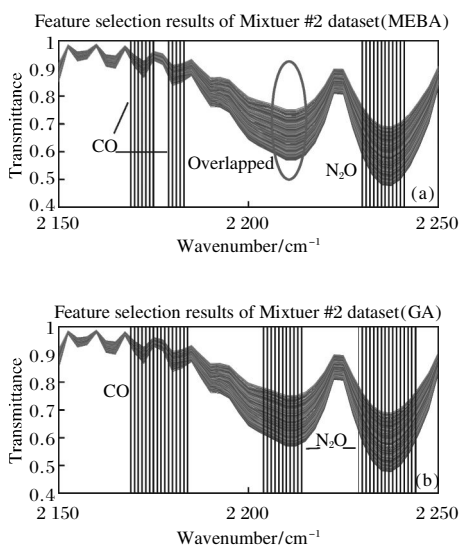


图 3 Mixture #2 数据库的特征选择结果
Fig. 3 Feature selection results of Mixture #2 dataset

3.2 ELM 与 PLS 浓度反演模型泛化性能对比

定量模型的建立方法是影响 MEBA 性能的重要因素,因此有必要对文中选择的 ELM 模型的泛化性能进行分析,这里将 ELM 与传统的 PLS 模型进行对比。针对每种模型,分别利用全光谱与筛选出的特征波长点建立浓度反演模型。选择测试集的平均相对误差、均方根误差和决定系数 R² 作为衡量模型

泛化性能的标准。经过主成分分析(PCA),Mixture #1 数据库的 PLS 浓度反演模型的主成分个数与 ELM 浓度反演模型的隐含层神经元个数均被设置为 8; Mixture #2 数据库的 PLS 浓度反演模型的主成分个数与 ELM 浓度反演模型的隐含层神经元个数均被设置为 10。Mixture #1 与 Mixture #2 数据库的测试集平均相对误差分别如图 4 和图 5 所示,对应的均方根误差和 R² 如表 1 所列。

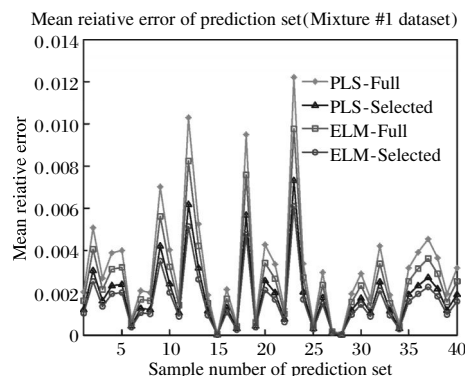


图 4 四种模型测试集平均相对误差(Mixture #1 数据库)

Fig. 4 Mean relative error of prediction set of four models(Mixture #1 dataset)

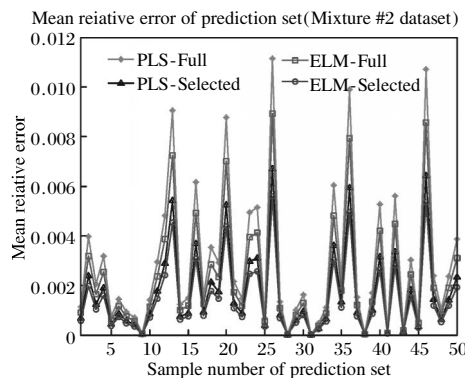


图 5 四种模型测试集平均相对误差(Mixture #2 数据库)

Fig. 5 Mean relative error of prediction set of four models(Mixture #2 dataset)

表 1 四种模型预测误差对比

Tab.1 Comparison of prediction error of four models

Model	Mixture #1database		Mixture #2database	
	RMSEP	R ²	RMSEP	R ²
ELM-Full	0.171	0.928	0.169	0.885
ELM-Selected	0.159	0.955	0.155	0.908
PLS-Full	0.238	0.776	0.229	0.754
PLS-Selected	0.193	0.901	0.189	0.868

Note:“-Full”表示利用全光谱建模;“-Selected”表示利用筛选出的特征波长点建模。

从图 4,图 5 及表 1 中不难发现:(1)利用 MEBA 筛选出的特征波长点建立的浓度反演模型的测试集平均相对误差显著小于利用全光谱波长点建立的浓度反演模型,这一方面表明,MEBA 可以有效地筛选出 CO 与 N₂O 的特征峰;另一方面也印证了建模前特征选择的必要性与重要性。(2)ELM 模型的测试集预测误差小于 PLS 模型,表明当输入光谱与目标组分间存在非线性映射时,与 PLS 模型相比,利用 ELM 建立的定量模型泛化性能更佳。

3.3 MEBA 与 BA, GA 及 PSO 算法的收敛性能对比

为了评价 MEBA 的收敛性能,文中将之与 BA、GA 及 PSO 算法进行对比。收敛时的迭代次数及陷入局部极小的情况作为评价的指标。Mixture #2 数据库对应的四种算法的迭代进化过程如图 6 所示,从中可以看出:

(1)迭代次数方面,MEBA<BA<GA<PSO,表明 MEBA 的收敛速度比其他三种算法要快。

(2)最终寻优目标函数值方面,MEBA<GA<PSO<BA,表明 MEBA 的全局搜索能力优于其他三种算法。

(3)陷入局部极小方面,MEBA<BA=GA=PSO,表明 MEBA 不易陷入局部极小。究其原因,是由于 MEBA 充分利用了 MEC 的异化操作以保证种群的多样性,使得在整个解空间中搜索的广度与深度增加,从而导致陷入局部极小的可能性降低。

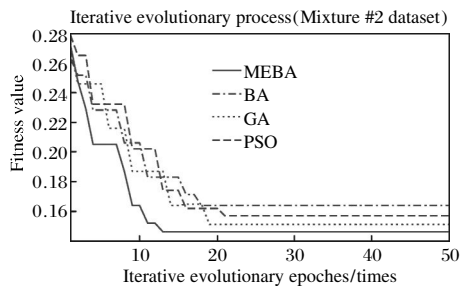


图 6 四种算法的迭代进化过程(Mixture #2 数据库)

Fig.6 Iterative evolutionary process of four algorithms(Mixture #2 dataset)

3.4 MEBA and PGSO 综合性能对比

除了 MEC,还有一些包含并行机制特点的算法见诸报道,如并行萤火虫群优化算法(PGSO)^[10]等。文中从以下四个方面对 MEBA 和 PGSO 的综合性能进行对比分析:迭代次数,筛选出的特征波长点个数, RMSEP 和 R^2 。另外,考虑到 UVE 算法可以快速剔除

除无信息变量,针对 MEBA 和 PGSO,均按是否结合 UVE 预处理进行讨论。实验结果如表 2 所列示。

表 2 MEBA 与 PGSO 综合性能对比

Tab.2 Comprehensive performance comparisons of MEBA and PGSO

Database	Method	M	N	RMSEP	R^2
Mixture #1 database	MEBA	16	88	0.159	0.955
	UVE-MEBA	12	82	0.161	0.936
	PGSO	21	101	0.186	0.924
Mixture #2 database	UVE-PGSO	14	83	0.162	0.933
	MEBA	11	58	0.155	0.908
	UVE-MEBA	8	55	0.164	0.896
	PGSO	16	73	0.184	0.872
	UVE-PGSO	12	66	0.168	0.887

从表 2 可以看出:(1)与 PGSO 算法相比,MEBA 的迭代次数更少,表明 MEBA 的收敛速度更快。同时,利用 MEBA 筛选出的特征波长点建立的浓度反演模型的预测误差更小,意味着泛化性能更优。这一现象的背后原因在于,PGSO 没有一个有效的机制来保证种群的多样性,相反,MEBA 充分利用了 MEC 的异化操作,使得整个搜索过程可以遍历整个解空间。(2)在利用 UVE 方法进行预处理,剔除掉无信息变量后,MEBA 和 PGSO 算法的迭代次数与筛选出的特征波长点个数均显著减小,表明在特征选择前,进行一轮“粗”筛选的效果更好。(3)MEBA 与 UVE-PGSO 的综合性能相当,也就是说,即使在不经 UVE 预处理剔除无信息变量的情况下,MEBA 仍然可以寻找到最佳的特征波长点组合。

4 结论

提出了一种新型的红外光谱特征选择方法。该方法充分利用思维进化计算的并行机制来实现快速搜索;利用思维进化计算的异化操作保证种群的多样性,即扩大整个解空间的搜索范围,提升全局搜索能力;利用蝙蝠算法的局部搜索能力,来获得局部范围的最优解。在此基础上,利用实验采集的两组混合气体数据库:Mixture #1(CO 与 CO₂ 混合)与 Mixture #2(CO、N₂O 与 CO₂ 混合)对思维进化蝙蝠算法的性能进行评价。同时,从收敛速度与搜索到的特征峰两个方面,与基本蝙蝠算法、遗传算法、粒子群优化及

并行萤火虫群优化算法等进行了对比。

实验结果表明:与传统方法相比,思维进化蝙蝠算法的收敛速度更快,陷入局部极小的概率更小。同时,在 Mixture #2 数据库中,CO 与 N₂O 的特征峰相互重叠的情况下,思维进化蝙蝠算法可以有效消除重叠光谱及干扰光谱的影响,准确地搜索到目标组分的特征峰,即 CO 的特征峰范围包括 2 090~2 110 cm⁻¹ 和 2 115~2 125 cm⁻¹,共包含 32 个波长点;N₂O 的特征峰范围为 2 225~2 250 cm⁻¹,共包含 26 个波长点。利用筛选出的特征波长点建立的浓度反演模型,测试集均方根误差为 0.155,决定系数可达 0.908,泛化性能显著优于传统方法。因此,思维进化蝙蝠算法作为新的特征选择手段,给混合气体乃至其他物质的红外光谱的特征筛选提供了有效参考。

参考文献:

- [1] Hong Mingjian, Wen Quan, Wen Zhiyu. New near infrared wavel -ength 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,官会丽.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] Chen Zhong, Zhang Liqing, Liu Heyang, et al. Successive projections algorithm and its application to selecting the wheat near-infrared spectral variables [J]. *Spectroscopy and Spectral Analysis*, 2010, 30(4): 949-952. (in Chinese)
成忠,张立庆,刘赫扬,等.连续投影算法及其在小麦近红外光谱波长选择中的应用 [J]. *光谱学与光谱分析*, 2010, 30(4): 949-952.
- [5] Zou Xiaobo, Zhao Jiewen, Povey M J W, et al. Variables selection methods in near-infrared spectroscopy[J]. *Analytica Chimica Acta*, 2010, 667: 14-32.
- [6] Yang Xinshe. A new metaheuristic bat-inspired algorithm[J]. *Nature Inspired Cooperative Strategies for Optimization*, 2010, 284: 65-74.
- [7] Wang Chuanlong, Xie Keming. Convergence of a new evolutionary computing algorithm in continuous state space [J]. *International Journal of Computer Math*, 2002, 79(1): 27-37.
- [8] Mirjalili S, Lewis A. S-shaped versus V-shaped transfer functions for binary particle swarm optimization [J]. *Swarm and Evolutionary Computation*, 2013, 9(1): 1-14.
- [9] Huang Guangbin, Zhu Qinyu, Siew C K. Extreme learning machine: theory and applications[J]. *Neurocomputing*, 2006, 70(1): 489-501.
- [10] Liu Jiakun. High dimension artificial glowworm swarm optimization algorithm analysis application and research [D]. Nanning: Guangxi University for Nationalities, 2012. (in Chinese)
刘佳昆.高维人工萤火虫群优化算法应用研究 [D]. 南宁:广西民族大学,2012.