

doi:10.3969/j.issn.1673-9833.2021.05.008

汽油辛烷值损失优化方案的数学建模与求解

陆天浩¹, 李玲玲¹, 陈宇峰¹, 汤琼²

(1. 湖南工业大学 计算机学院, 湖南 株洲 412007; 2. 湖南工业大学 理学院, 湖南 株洲 412007)

摘要: 辛烷值作为汽油的商品牌号, 是反映汽油燃烧性能的最重要指标。针对有效利用重油资源问题, 以辛烷值作为评价指标, 使用线性插值和拉伊达准则等方法对数据集进行预处理; 通过主成分分析与层次聚类, 寻求主要变量进行降维处理; 通过建立基于BP神经网络的辛烷值损失预测模型, 对辛烷值及其指标进行预测; 最后使用遗传算法迭代得到最优参数集合, 并可视化展示变化轨迹。经过实验, 预测模型迭代到2 048次时, 模型损失为 9.9826×10^{-6} 。从数据集筛选出硫含量不大于 $5 \mu\text{g/g}$ 的样本和在取值范围内变化的主要操作变量, 通过主要操作变量的取值范围和步长的约束, 找到满足辛烷值损失大于30%的样本, 即对应主要变量优化的操作条件。研究表明, 采用的数学模型与优化算法可以较好地对比辛烷值损失问题进行预测与优化。

关键词: 汽油; 辛烷值; 主成分分析; 层次聚类; 神经网络; 遗传算法

中图分类号: O29

文献标志码: A

文章编号: 1673-9833(2021)05-0062-08

引文格式: 陆天浩, 李玲玲, 陈宇峰, 等. 汽油辛烷值损失优化方案的数学建模与求解[J]. 湖南工业大学学报, 2021, 35(5): 62-69.

Mathematical Modeling and Solution of Optimal Scheme for Gasoline Exquisite Octane Loss

LU Tianhao¹, LI Lingling¹, CHEN Yufeng¹, TANG Qiong²

(1. College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China;

2. College of Science, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: As a trademark of gasoline, octane number is the most important indicator reflecting the combustion performance of gasoline. In view of the flaw of effective utilization of heavy oil resources, with octane as an evaluation index, linear interpolation and Laida criterion are used to preprocess the data set, meanwhile principal component analysis and hierarchical clustering are adopted to obtain the main variables for dimension reduction. An octane number loss prediction model based on BP neural network has been established for the prediction of the octane number and its indicators. Finally, the genetic algorithm is used to iteratively obtain the optimal parameter set, with the change trajectory displayed visually. Based on the experimental results, with the number of iterations of the prediction model being 2 048, the loss of the model is maintained at 9.9826×10^{-6} . The sulfur content selected from the data set is no more than $5 \mu\text{g/g}$ and the main operating variables of the samples are unchanged within the value range. According to the constraints of the value range and step size of the main operating variables, the optimal operating conditions

收稿日期: 2020-12-18

作者简介: 陆天浩(1995-), 男, 宁夏银川人, 湖南工业大学硕士生, 主要研究方向为计算机视觉与模式识别,
E-mail: 18813093350@163.com

通信作者: 汤琼(1972-), 女, 湖南浏阳人, 湖南工业大学教授, 硕士生导师, 主要从事计算数学方面的教学与研究,
E-mail: zgzxysx@163.com

corresponding to the main variables are found to meet an octane loss of more than 30%. The research results show that the adopted mathematical model and optimization algorithm can be used for the prediction and optimization of the octane number loss problem.

Keywords: gasoline; octane number; principal component analysis; hierarchical clustering; neural network; genetic algorithm

0 引言

汽油的辛烷值作为汽油的商品牌号(例如 89#、92#、95#),是衡量汽油发动机燃料抗爆性能优劣的重要指标。辛烷值越高,表示其抗爆性能越好,发动机压缩比越高^[1]。辛烷值的提高可以提升发动机功率,增加车辆行程;同时还能起到节约燃料,减轻使用者经济负担的作用。对于小型车辆而言,汽油是其主要的燃料,汽油燃烧产生的尾气是影响大气环境的重要因素。辛烷值反映了汽油的燃烧性能,然而当前的脱硫催化裂化汽油技术,使汽油的辛烷值大大降低。因此提高辛烷值是提高经济效益和改善环境的一个重要手段^[2]。

为了解决上述问题,针对“华为杯”第十七届中国研究生数学建模竞赛^①B题:汽油辛烷值建模,将BP(back propagation)神经网络模型与遗传算法^[3-8]引入预测与优化任务中,并通过预测优化找到符合条件的优化方案。

1 数据处理

1.1 数据预处理

用数学建模竞赛B题所提供的数据集作为原始数据集。在数据集中,共有325个样本,根据样本采集时间排序,采集时间为2017-04—2020-05。每个样本包含由原料性质、产品性质、待生吸附剂性质、再生吸附剂性质等不同操作变量构成的354个采集位节点。由于是原始数据,不同样本中的不同位节点存在未采集数据或数据缺失等情况。如果不同样本的同一位节点缺失数据较多,则该位节点不能准确反映变化情况,需要删除该位节点。如果不同位节点的同一样本数据较多,同样需要删除该样本。对于缺失值个数较少的位节点,可根据临近时间的样本对应位节点值进行填充。再对筛选后的数据采用 3σ 准则去除异常值;根据实际应用背景确定取值范围,将超过范围的位节点或对应样本数据剔除。具体数据处理

^①本文为“华为杯”第十七届中国研究生数学建模竞赛二等奖论文。

如下。

1) 遍历整个数据表,将采集值缺失过多的样本及位节点删除。经过筛选,共删除12个位节点,具体名称如表1所示。这类位节点由于信息缺失过多,因此不能清晰体现该位节点对辛烷值损失的影响。

表1 第一步被删除的位节点
Table 1 Removed sites in the first step

序号	位节点名	序号	位节点名
1	S-ZORB.FT_1501.PV	7	S-ZORB.FC_1104.DACA
2	S-ZORB.FT_1002.PV	8	S-ZORB.FT_2803.DACA
3	S-ZORB.FC_1202.PV	9	S-ZORB.FT_1502.DAC
4	S-ZORB.FT_1501.TOTAL	10	S-ZORB.FC_2432.PIDA.SP
5	S-ZORB.FT_5102.P	11	S-ZORB.TEX_3103A.DAC
6	S-ZORB.FT_2901.DACA	12	S-ZORB.FT_5102.DACA.PV

2) 对位节点进行过滤操作,删除数值全部为空值或无法通过相邻样本进行填充的位节点。前者没有数值变化,不能体现该位节点对辛烷值损失的影响。后者只有部分时间段样本,不能完整地体现该位节点对辛烷值损失的作用。过滤后,删除19个位节点,具体名称如表2所示。

表2 第二步被删除的位节点
Table 2 Removed sites in the second step

序号	位节点名	序号	位节点名
1	S-ZORB.FC_2301.PV	11	S-ZORB.FT_2901.DACA
2	S-ZORB.FT_1501.PV	12	S-ZORB.FT_2002.DACA
3	S-ZORB.FT_5104.PV	13	S-ZORB.FC_1104.DACA
4	S-ZORB.FT_9101.PV	14	S-ZORB.FT_2803.DACA
5	S-ZORB.FT_1002.PV	15	S-ZORB.FT_1502.DACA
6	S-ZORB.FC_1202.PV	16	S-ZORB.FT_3702.DACA
7	S-ZORB.FC_3103.PV	17	S-ZORB.TEX_3103A.DACA
8	S-ZORB.FT_1002.TOTAL	18	S-ZORB.FT_5102.DACA.PV
9	S-ZORB.FT_1501.TOTAL	19	S-ZORB.FT_5204.DACA.PV
10	S-ZORB.FT_510.PV		

3) 对于存在少量空缺值的位节点,用相邻样本的均值进行填补,需要补充的位节点如表3所示。

4) 根据不同操作变量实际取值范围,剔除位节点采集值异常的32,29,27号样本。

5) 使用 3σ 准则去除异常值及其所带来的影响。

完成上述预处理后,将285,313号样本结果加入到对应位置。

表3 缺省值填补位节点

Table 3 Nodes filled by the default value

序号	位节点名	序号	位节点名
1	S-ZORB.FT_1204.DACA.PV	4	S-ZORB.FC_2432.DACA
2	S-ZORB.FC_2432.PIDA.SP	5	S-ZORB.FT_1204.PV
3	S-ZORB.FT_2431.DACA		

1.2 数据降维

模型1 主成分分析

主成分分析法是使用线性降维的方法来寻找模型的主要操作变量,以达到精简影响因素数量的目的。本文使用SPSS进行主成分分析,使多维的数据降维得到少数几个综合指标,具体操作步骤如下:

1) 数据标准化

设 x_1, x_2, \dots, x_m 为 m 个主成分分析变量, a_{ij} 为 m 个评价对象中第 i 个评价对象对应于第 j 个分析变量的取值。将每个 a_{ij} 值标准化,

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j} (i, j = 1, 2, \dots, m), \quad (1)$$

式中: $u_j = \frac{1}{m} \sum_{i=1}^m a_{ij}$;

$$s_j = \frac{1}{m} \sum_{i=1}^m (a_{ij} - u_j)^2。$$

标准化指标变量,即

$$\tilde{x}_j = \frac{x_j - \mu_j}{s_j} (j = 1, 2, \dots, m)。 \quad (2)$$

2) 计算数据的相关系数矩阵

相关系数矩阵 R 的具体表达式为

$$R = (r_{pq})_{m \times m}, \quad (3)$$

式中, r_{pq} 为第 p 个指标与第 q 个指标之间的相关系数,且

$$r_{pq} = \frac{\sum_{k=1}^n (\tilde{a}_{kp} \cdot \tilde{a}_{kq})}{n-1} (p, q = 1, 2, \dots, m), \quad (4)$$

$$r_{pp} = 1, r_{pq} = r_{qp}。$$

使用主成分分析法对数据集中的320个位节点进行分析。第一次实验中,将主成分个数设置为25,共筛选出200多个变量。但由于决定主成分因子的操作变量种类各不相同,在工艺和设备操作上没有很强的关联性。因此将主成分个数设为5,然后重新实验,结果如下:反应过滤器压差,反应器上部温度, SZORB.FT_9001.TOTAL, SZORB.FT_5201.TOTAL, SZORB.FT_9102.TOTAL。第二次实验结果优于第一次,但由于传统方法的关联模型变量少、要求高,因此响应慢、效果差,且5个主成分因子只能解释59.374%的结果。

总的来说,通过对位节点进行降维,可以得到一部分主要操作变量。但是由于操作变量之间具有高度非线性和相互强耦合的关系,使用主成分分析法筛选出的主要操作变量影响因子较低,不具有很强的解释性,因此本文采用非线性的降维方法。

模型2 层次聚类法

由于数据集中的320个有效位节点之间具有高度非线性和相互强耦合关系,为了寻找具有代表性和独立性的主要操作变量,再采用非线性的层次聚类法进行筛选。最后将主成分分析与层次聚类法筛选出的结果进行综合分析,得出最终建模需要的主要操作变量。构建过程如下:

参考欧几里德距离评价,在数据不规范时,皮尔逊相关系数会给出更好的结果,故本文使用皮尔逊相关系数来寻找模型的主要操作变量。具体算法见式(5)(6)。

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{E[(X - E(X))(Y - E(Y))]}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{\sum_{i=1}^n [(X_i - \mu_X)(Y_i - \mu_Y)]}{\sqrt{\sum_{i=1}^n (X_i - \mu_X)^2} \sqrt{\sum_{i=1}^n (Y_i - \mu_Y)^2}}, \quad (5)$$

$$\text{dist}(X, Y) = 1 - \rho_{XY}。 \quad (6)$$

式(5)(6)中: X, Y 为两个不同的变量;

μ_X, μ_Y 和 $D(X), D(Y)$ 分别为变量 X, Y 的均值和方差;

ρ_{XY} 为两变量 X, Y 间的皮尔逊相关系数,当2个变量完全匹配时, $\rho_{XY}=1$,当2个变量毫无关系时, $\rho_{XY}=0$ 。

为了使相似度越大的两个变量之间距离越小,采用1与皮尔逊相关系数的差值来衡量^[9],见式(6)。

使用SPSS进行层次聚类分析,再与主成分分析得到的结果综合分析后,得到了13个主要变量:氢油比、反应过滤器压差、反应器上部温度、反应器底部温度、烯烃含量、芳烃含量、反应器顶低压差、待生吸附剂性质(焦炭含量)、待生吸附剂性质(硫含量)、再生吸附剂性质(焦炭含量)、再生吸附剂性质(硫含量)、辛烷值和硫含量。结合工艺要求和操作经验发现这13个主要变量符合实际情况,并在后面的辛烷值损失模型建立和优化过程中,可呈现较好的效果,证明了这13个主要变量具有有效性和代表性。

2 基于 BP 神经网络的辛烷值损失预测模型

反向传播网络可以拟合任何非线性函数, 是可以实现从输入到输出的端到端网络, 而且训练网络的方法非常简单、有效, 只需要学习网络参数即可。通过数据预处理与数据降维, 将 13 个主要变量作为神经网络的输入进行训练。基于 BP 神经网络的辛烷值损失预测模型网络结构图如图 1 所示^[8]。神经网络模型可分为两部分: 第一部分神经网络有 800 层, 作为特征提取层; 另一部分为单层全连接层, 作为预测层。需要注意的是, 在模型选择时, 输入输出的维数固定, 中间的隐藏层部分需要考虑是否全连接。图中的箭头表示数据流动的方向, 即模型的流程; w 表示神经网络中的权重矩阵, 权重可通过训练得到; B 表示各层神经元的偏置, 可以使模型具有更鲁棒的表达能力。

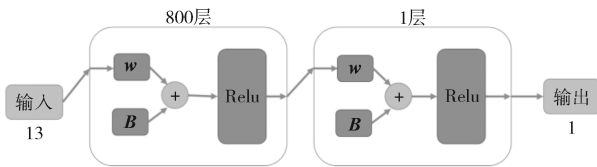


图 1 基于 BP 神经网络的辛烷值损失预测模型网络结构图
Fig. 1 Network structure diagram of octane loss prediction model based on BP neural network

模型 (反向传播梯度) 在更新网络参数时, 为避免波动过大出现梯度消失或梯度爆炸现象, 导致模型崩溃, 需要对输入进行零均值化与归一化, 使输入数据的分布统一。经过归一化, 输入被约束在 $[-1, 1]$ 的区间内。

第一步零均值化

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i, \quad (7)$$

式中: x_i 为某一主要变量的某条样本采样值。

m 为主要变量的样本个数。

将降维后的数据进行 $\bar{X} = X - \mu$ 变换, 从而使训练集分布中心与零点重合, 完成零均值化^[3]。

第二步归一化方差

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu)^2. \quad (8)$$

将零均值化的数据除以 σ^2 , 使所有数据用 $[0, 1]$ 内的数表示。经过零均值化和归一化后, 神经网络的代价函数更平滑、易于优化^[4]。

在神经网络中添加激活函数, 可以增加网络的非线性表达能力, 并解决线性模型不能很好描述的函数

关系^[5]。神经网络中所采用 ReLU 激活函数为

$$f(z) = \max(0, w^T z + B), \quad (9)$$

式中: z 为激活函数的输入, 在本文的网络中为激活函数前一层神经元的输出;

B 为对应层神经元的偏置 (参见图 1)。

ReLU 函数在正区间解决了梯度消失的问题, 而且使神经网络的收敛速度和计算速度较使用其他激活函数 (如 sigmoid 等) 有明显提高, 起到了类似 Dropout 的各层神经元随机连接的效果。

考虑到神经网络的输入数据为筛选出的主要操作变量, 故采用均方误差 (mean square error, MSE) 来进行模型的评价。MSE 可以评价数据的变化程度, 其值越小说明模型预测结果具有更好的精确度^[6], 且具有较小的波动。具体表达式如下:

$$MSE = \frac{1}{m} \sum_{i=1}^m w_i (y_i - \hat{y}_i)^2. \quad (10)$$

式中: y_i 、 \hat{y}_i 分别为辛烷值损失真实值和模型预测值; w_i 为神经网络中各预测值的权重。

3 模型求解

在对神经网络模型的实际训练过程中, 采用 Adam 优化器进行梯度更新。相比于传统的随机梯度下降, Adam 通过计算梯度矩阵为各个参数设置更加灵活的学习速率, 对变化频繁的参数采用大步长进行学习, 对于稀疏的参数采用小步长进行学习。第 k 步梯度 g_k 更新如式 (11) 所示。

$$g_{k+1} = g_k - \frac{\alpha g_k}{\sqrt{\sum_{i=1}^k g_i^2}}, \quad (11)$$

式中 α 为调节系数, 用于调节变化快慢。

将整个数据集划分为 80% 的训练集和 20% 的测试集。测试集不参与模型训练, 在测试集上通过模型的预测结果来观察模型的实际泛化能力。经过 2 048 Epochs 训练, 模型的损失函数变化如图 2 所示。

由图 2 可知, 随着训练次数的增加, 在测试集上的损失值逐渐降低, 在 2 000 次左右达到了 9.928×10^{-6} , 这说明模型有良好的泛化性能。

以损失函数的变化作为本研究中神经网络预测模型的测试标准之一。使用 Matlab 软件建立网络的训练参数如下: 迭代次数为 2 048, 学习率为 0.01。训练时模型参数的变化如图 3 所示。由图 3a 可以看出, 随着 epoch 的增加, 模型梯度逐渐减小, 变化趋势逐渐稳定。图 3b 可以看出在训练时没有出现模型崩溃现象, 具有很好的稳定性。

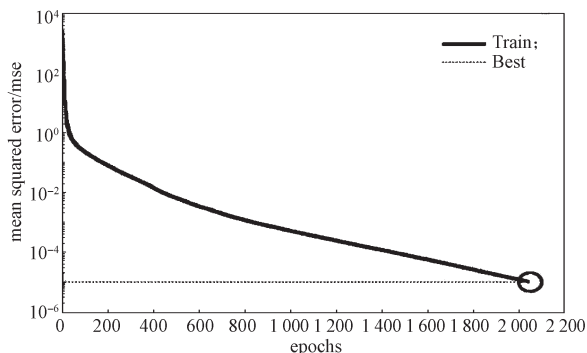
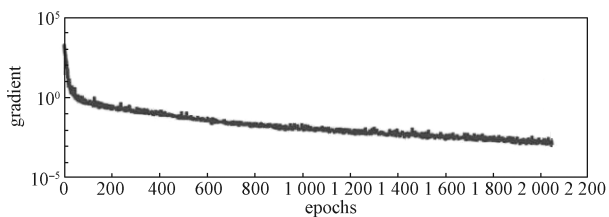
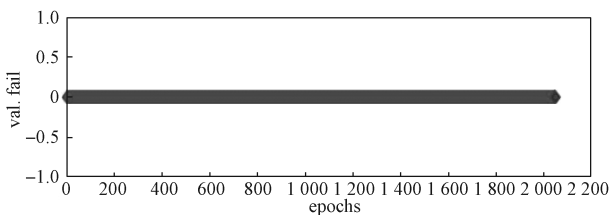


图2 损失函数的变化图

Fig. 2 Change diagram of the loss function



a) 反向传输梯度变化



b) 模型验证出错次数

图3 训练时神经网络模型参数变化图

Fig. 3 Neural network model parameter variation diagram during training

经过训练,训练模型的最终回归结果如图4所示。由图4可以看出,目标值和输出结果基本上在直线 $y=x$ 附近分布,说明训练结果比较理想。

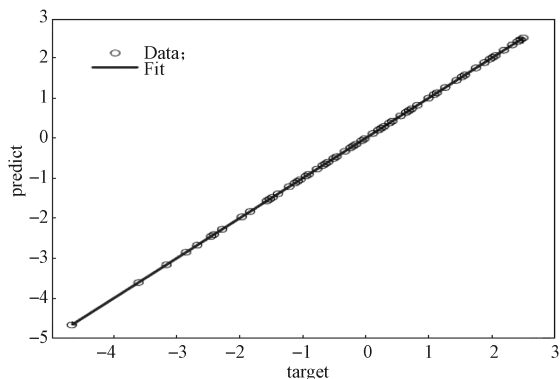


图4 模型的回归结果

Fig. 4 Regression results of the model

图5为各样本的辛烷值损失实际值与预测值的分布情况。由图5可以看出,模型拟合效果明显,可以较好地表达辛烷值的变化情况,并且能取得很好的预测效果。

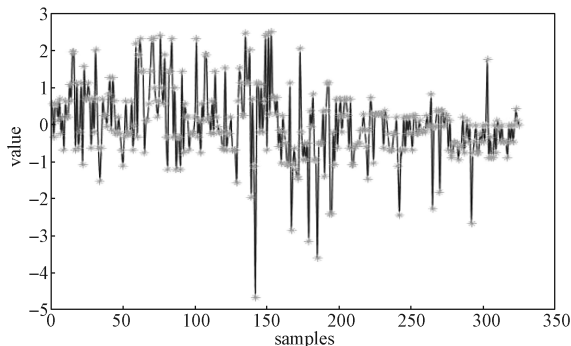


图5 各样本预测值与真实值分布图

Fig. 5 Predicted value and true value distribution of each individual sample

图6为3509次迭代的损失函数变化图。由图可以看出,在训练2000次以后,损失函数下降趋势基本趋于稳定,与图2中2048次迭代的结果相比,MSE在预测和训练上的变化区别不大,拟合程度较高,这说明模型具有良好的有效性和泛化能力。

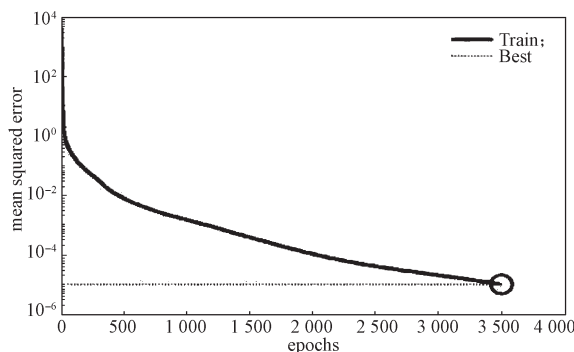


图6 3509次迭代损失函数变化图

Fig. 6 Loss function variation diagram for 3509 iterations

4 基于遗传算法的方案优化

遗传算法是一种随机化搜索方法,是一种模拟达尔文进化论和孟德尔遗传学机理的计算模型。遗传算法由编码、适应度评估和遗传运算3部分组成^[7]。根据损失散点图,将损失变化通过三角函数进行拟合,即可得到每一个样本的损失值。由于每一个样本包含13个主要操作变量,损失值可以由操作变量表达,在此可将问题转变为基于主要操作变量的曲线优化问题。利用遗传算法解决多元非线性回归问题,并向最小化损失值的方向优化,得到13个主要操作变量的对应取值。根据变量的取值范围,得到对应主要变量的优化操作条件。

对于本问题,使用传统方法难以获得最优解,而多峰最优化问题具有多个局部解的特性,因此使用多峰非线性方法来求解,具体实验参数设置如表4所示。其中,自变量上下界用来表示每个主要操作变量取值

范围, 交叉概率、变异概率用于控制个体改变频率, 设置迭代次数为 1 000, 使算法能够得到足够稳定的优化解。

表 4 实验参数设置

Table 4 Experimental parameter setting

代 码	意 义
gen=1 000	迭代次数
upper=[0.37, 35, 435, 430, 110, 34.67, 26.2, 12.15, 14.31, 8.92, 13.34, 91.7, 5]	自变量上界
lower=[0.20, 10, 410, 399, 50, 14.6, 16, 1.01, 2.94, 0.25, 1.43, 85.3, 3.2]	自变量下界
pc=0.25	交叉概率
pm=0.01	变异概率
pop_num=100	种群大小
delta=[0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1, 0.000 1]	编码精度
fun=0	0 最大化, 1 最小化

经过实验, 遗传算法的损失变化如图 7 所示。由图 7 可以看出, 随着迭代次数增加, 损失的波动幅度逐渐减小, 最终解逐渐趋于稳定。损失变化曲线表达式为

$$\begin{cases} \max f(x_1, x_2) = 21.5 + x_1 \sin(4\pi x_1) + \\ \quad x_2 \sin(20 + \pi + x_2); & (12) \\ \text{s.t. } -3.0 \leq x_1 \leq 12.1, 4.1 \leq x_2 \leq 5.8 \end{cases}$$

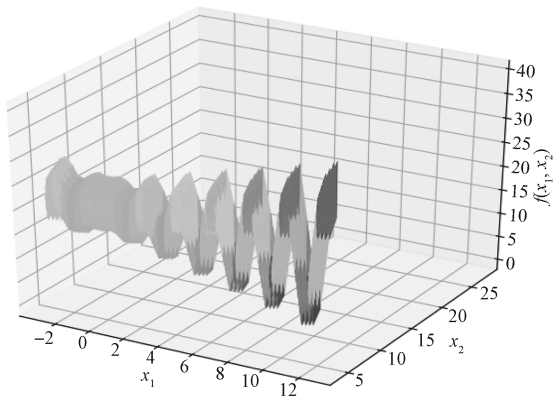


图 7 遗传算法的损失变化曲线

Fig. 7 Loss change curve of genetic algorithm

遗传算法最优个体得分曲线如图 8 所示。由图 8 可能看出, 遗传算法在 400 次迭代后每代最优函数值变化逐渐稳定, 这证明算法已逐渐收敛于最优解。

通过上述遗传算法, 可以得到一批子代, 每个子代对应一组变量。根据辛烷值损失值与 13 个变量之间的关系, 建立如下损失值与主要变量间的表达式:

$$Y=f(x_1, x_2, \dots, x_{13}). \quad (13)$$

由于优化问题一般是从凸函数最小化寻找最优解, 因此采取最小化 f 值的方式, 获得 x 变量的取值,

x 对应 13 个主要操作变量。经过遗传算法的编码解码, 从遗传算法中选择降幅大于 30% 的样本, 每个样本对应主要变量优化后的一种操作条件。符合条件的优化方案, 按照降幅比降序排列后, 结果如表 5 所示。当 13 个主要变量分别优化至表 5 首行对应数值时, 所能达到的降幅比为 44.822%, 是所有得到的优化方案中的最优解。

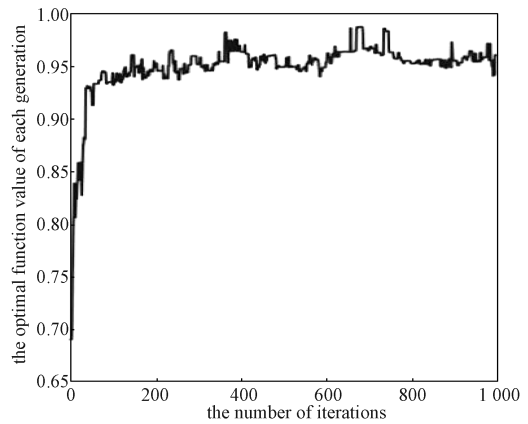


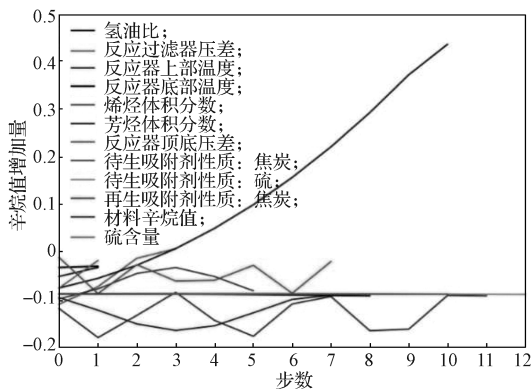
图 8 每代最优个体得分变化曲线

Fig. 8 Optimal individual score variation for each generation

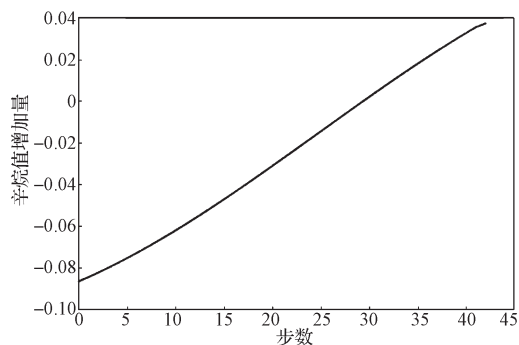
当某一变量逐步优化, 其他变量不变时, 辛烷值增加量与优化步数之间的变化情况如图 9a 和 9b 所示。由于再生吸附剂性质硫优化迭代步数较多, 因此通过图 9b 展示。由图 9 可以看出, 辛烷值增加量的变化与氢油比、反应器底部温度和反应器顶底压差变化呈正相关; 随着变量反应过滤器压差值越接近最优理想状态, 对辛烷值的影响变弱; 辛烷值增加量变化与反应器上部温度变化呈先负相关后正相关的关系, 且变化趋势大致相同, 而待生吸附剂中的焦炭含量对辛烷值变化影响相反; 辛烷值增加量与烯烃含量变化呈不断抖动的关系, 正负相关性交替出现; 辛烷值增加量与芳烃含量呈近似指数变化的正相关关系, 且越接近最优状态影响越明显; 再生吸附性质(焦炭含量)对辛烷值增加量产生不断交替的正负相关性, 但总体呈微弱的正相关性; 辛烷值增加量与再生吸附性质(硫含量)呈倒二次函数变化趋势; 原材料对产品辛烷值增加量的影响呈现出负相关的特性, 且随着其值逼近最优值, 对产品辛烷值增加量的负向影响力度愈加强烈。由图 9c 可知, 当所有主要变量同时进行优化时, 主要变量对辛烷值增加量的综合影响可以分为两个阶段, 从一开始的初始状态到第 10 次迭代, 对辛烷值增加量的变化影响明显; 第 10 次迭代之后, 对辛烷值的影响虽然仍然是正相关, 但其影响程度下降了很多, 其中一个原因是有些变量已经达到最优解。

表5 符合条件的优化方案
Table 5 Satisfactory optimized schemes

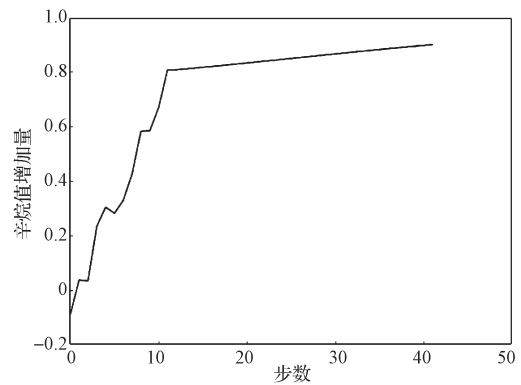
氢油比	反应过滤器压差 /kPa	反应器温度 /℃		烯烃体积分数 /%	芳烃体积分数 /%	反应器顶底压差 /kPa	待生吸附剂性质		再生吸附剂性质		辛烷值	硫含量 / (μg·g ⁻¹)	降幅比
		上部	底部				焦炭质量分数 /%	硫质量分数 /%	硫质量分数 /%	焦炭质量分数 /%			
0.321 085	26.148 91	419.122 7	419.023 4	75.644 52	34.251 82	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	88.966 66	4.248 842	0.448 217
0.320 420	26.148 91	418.732 1	419.038 5	75.644 52	34.251 98	19.295 38	8.435 76	9.450 353	5.217 203	2.422 447	88.966 66	4.477 364	0.365 672
0.321 749	26.145 86	418.732 1	419.023 4	73.710 92	34.251 98	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	89.066 66	4.248 842	0.365 672
0.320 420	26.145 86	418.732 1	419.015 8	73.710 92	34.251 98	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	89.066 66	4.248 842	0.365 672
0.321 085	26.145 86	418.732 1	419.023 4	60.585 91	34.251 82	19.295 38	8.437 12	9.450 353	5.217 203	2.422 447	89.066 66	4.248 842	0.365 672
0.320 420	26.148 91	418.732 1	419.030 9	60.644 50	34.251 82	19.295 38	8.437 21	9.450 353	5.217 203	2.422 447	89.066 66	4.248 842	0.365 672
0.319 091	26.148 91	418.732 1	419.015 8	60.585 91	34.251 98	19.295 38	8.437 12	9.461 457	5.217 203	2.422 447	89.066 66	4.252 358	0.362 959
0.320 420	26.148 91	418.732 1	419.030 9	60.644 50	34.251 82	19.295 38	8.437 12	9.450 353	5.217 203	2.422 447	89.066 66	4.248 842	0.355 251
0.321 085	27.708 37	417.169 6	419.023 4	75.644 52	34.251 82	19.295 38	8.437 12	9.450 353	5.217 203	2.422 447	89.066 66	4.248 842	0.355 251
0.320420	26.145 86	418.732 1	419.023 4	73.710 92	34.251 98	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	89.066 66	4.248 842	0.355 251
0.318 427	26.148 91	418.732 1	419.023 4	75.585 92	34.251 98	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	89.066 66	4.248 842	0.352 978
0.320 420	26.148 91	418.732 1	419.030 9	60.585 91	34.251 82	19.292 89	8.437 12	9.405 353	7.655 659	2.422 447	89.066 66	4.248 842	0.329 193
0.320 420	26.148 91	418.732 1	419.030 9	58.710 91	34.251 98	19.295 38	8.437 12	9.461 457	5.217 203	2.422 447	89.066 66	4.252 358	0.329 193
0.321 085	26.148 91	418.732 1	419.023 4	75.644 52	34.251 98	19.611 65	8.437 12	12.659 290	5.217 203	2.422 447	89.066 66	4.248 842	0.327 694
0.320 420	26.148 91	418.732 1	419.030 9	60.585 91	34.251 82	19.292 89	8.437 12	9.805 668	7.655 659	2.422 447	89.066 66	4.248 842	0.327 694
0.320 420	26.148 91	418.732 1	419.030 9	60.644 50	34.251 82	19.295 38	8.437 12	9.405 053	5.217 203	2.422 447	89.066 66	4.248 842	0.318 811
0.321 749	26.145 86	418.732 1	419.015 8	60.585 91	34.251 98	19.295 38	8.435 76	9.816 772	7.655 659	2.422 447	89.066 66	4.477 364	0.318 811
0.320 420	26.148 91	418.732 1	419.023 4	75.644 52	34.251 98	19.295 38	8.435 76	9.405 053	5.217 203	2.422 447	88.966 66	4.477 364	0.318 191
0.318 427	26.148 91	418.732 1	419.015 8	60.585 91	34.251 98	19.295 38	8.435 76	9.816 772	7.655 659	2.422 447	89.066 66	4.477 364	0.316 634
0.299 824	27.708 37	417.145 1	419.023 4	75.585 92	34.251 98	19.295 38	8.437 12	9.095 038	5.217 203	2.422 447	88.966 66	4.248 842	0.314 446
0.320 420	26.148 91	418.732 1	419.030 9	60.585 91	34.251 82	19.292 89	8.437 12	9.405 053	7.655 659	2.422 447	89.066 66	4.248 842	0.313 990
0.320 420	27.711 42	418.732 1	419.030 9	60.644 50	34.251 82	19.295 38	8.437 12	9.405 053	5.217 203	2.422 447	89.066 66	4.248 842	0.312 918
0.320 420	26.148 91	418.732 1	419.030 9	60.644 50	34.251 98	19.292 89	8.437 12	9.805 668	5.217 203	2.422 447	89.066 66	4.248 842	0.311 814
0.320 420	26.148 91	418.732 1	419.030 9	60.644 50	34.251 82	19.295 38	8.437 12	9.405 053	5.217 203	2.422 447	89.066 66	4.248 842	0.309 196
0.321 749	26.145 86	418.732 1	419.015 8	60.703 10	34.251 98	19.614 14	8.435 76	9.816 772	7.655 659	2.422 447	89.066 66	4.477 364	0.301 786



a) 各主要变量单独变化分别对辛烷值增量的影响



b) 变量再生吸附剂性质硫变化对辛烷值增量的影响



c) 各变量综合变化对辛烷值增量的影响

图9 各主要变量变化及辛烷值增量变化轨迹图

Fig. 9 Trajectories of major variables with their octane numbers

5 结语

本文通过层次聚类的方法，对预处理后的原数据集进行降维处理，建立基于BP神经网络的辛烷值损失预测模型，对辛烷值损失及其指标进行预测；使用遗传算法迭代得出辛烷值损失降幅比大于30%的优化方案及最优优化策略，并通过可视化将各个操作

变量的变化轨迹展示出来加以分析。实验结果表明,使用非线性层次聚类法比使用线性聚类法效果更好,能够很好地对数据集进行降维,降低了神经网络模型的训练与收敛难度。预测模型经过 2 048 次迭代后,神经网络模型的损失函数下降趋势基本稳定。在数据集上,采用 8:2 的比例构建训练集和测试集,并在测试集上达到了 $9.982\ 6 \times 10^{-6}$ 的误差精度,能满足模型预测的要求和精度。根据样本硫含量不大于 $5\ \mu\text{g/g}$ 、主要操作变量的取值范围,以及 13 个主要变量的变化步长,利用遗传算法迭代得到最优的参数集合。从集中找到子代满足辛烷值损失降幅比达到 30% 的样本并排序,得到全部优化方案以及最优优化方案。在最优方案中辛烷值损失的降幅比达到 44.822%,且各主要变量最优值均处于合理取值范围内。结果表明,本文所采用的数学模型与优化算法可以很好地对辛烷值损失问题进行预测与优化。

参考文献:

- [1] 李宏涛, 魏文平, 蒋国权, 等. 汽油加氢脱硫与辛烷值损失的平衡 [J]. 石化技术与应用, 2017, 35(4): 326-328.
LI Hongtao, WEI Wenping, JIANG Guoquan, et al. Equilibrium Between Gasoline Hydrodesulfurization and Octane Loss[J]. Petrochemical Technology and Application, 2017, 35(4): 326-328.
- [2] 练雄辉, 刘冰. 提高催化裂化汽油辛烷值措施探讨 [J]. 化工管理, 2019(6): 165-166.
LIAN Xionghui, LIU Bing. Discussion on Measures to Increase the Octane Number of FCC Gasoline[J]. Chemical Enterprise Management, 2019(6): 165-166.
- [3] 王爱丽, 薛冬, 吴海滨, 等. 基于条件生成对抗网络的手写数字识别 [J]. 液晶与显示, 2020, 35(12): 1284-1290.
WANG Aili, XUE Dong, WU Haibin, et al. Handwritten Digit Recognition Based on Conditional Generation Adversarial Network[J]. Chinese Journal of Liquid Crystals and Displays, 2020, 35(12): 1284-1290.
- [4] 极客 Array. 吴恩达深度学习笔记 (35): 加快归一化的网络训练方法 [EB/OL]. (2018-12-21) [2020-09-20]. https://blog.csdn.net/harpoon_fly/article/details/85144763.
Geek Array. Wu Enda's Deep Learning Notes-Speeding up Normalized Network Training Method[EB/OL]. (2018-12-21) [2020-09-20]. https://blog.csdn.net/harpoon_fly/article/details/85144763
- [5] XU B, WANG N Y, CHEN T Q, et al. Empirical Evaluation of Rectified Activations in Convolutional Network[EB/OL]. [2020-09-20]. https://www.researchgate.net/publication/275974753_Empirical_Evaluation_of_Rectified_Activations_in_Convolutional_Network.
- [6] 苏美红. 机器学习中损失函数问题研究 [D]. 西安: 西北大学, 2015.
SU Meihong. Research on Loss Function in Machine Learning[D]. Xi'an: Northwest University, 2015.
- [7] 柏熙, 林桂玲, 李成文, 等. 基于神经网络和遗传算法的智能监控系统的数学建模 [J]. 矿业安全与环保, 2006, 33(1): 15-17.
BAI Xi, LIN Guiling, LI Chengwen, et al. Mathematical Modeling of Intelligent Monitoring System Based on Neural Network and Genetic Algorithm[J]. Mining Safety and Environmental Protection, 2006, 33(1): 15-17.
- [8] 李尚志. 数学建模竞赛教程 [M]. 南京: 江苏教育出版社, 1996: 314-331.
LI Shangzhi. Mathematical Modeling Contest Course[M]. Nanjing: Jiangsu Education Press, 1996: 314-331.
- [9] 姜启源. 数学模型 [M]. 北京: 高等教育出版社, 1987: 305-335.
JIANG Qiyuan. Mathematical Model[M]. Beijing: Higher Education Press, 1987: 305-335.

(责任编辑: 邓光辉)