DOI:10. 3969/j. issn. 2097-0706. 2023. 03. 002

# 基于PCA与SSA-LightGBM的油浸式变压器故障诊断方法

Oil-immersed transformer fault diagnosis method based on PCA and SSA-LightGBM

肖宏磊[1],留毅[2],夏红军[1],缪宇峰[2],俞啸玲[1],杨海琦[3*]
XIAO Honglei[1],LIU Yi[2],XIA Hongjun[1],MIAO Yufeng[2],YU Xiaoling[1],YANG Haiqi[3*]

(1.国网浙江省电力有限公司杭州市余杭区供电公司,杭州 311100;2.杭州电力设备制造有限公司余杭群力成套电气制造分公司,杭州 311100;3.东北电力大学 机械工程学院,吉林 吉林 132012)
(1.Hangzhou Yuhang Power Supply Company,State Grid Zhejiang Power Company Limited,Hangzhou 311100,China;2.Yuhang Qunli Complete Sets Electricity Manufacturing Branch of Hangzhou Electric Power Equipment Manufacturing Company Limited,Hangzhou 311100,China;3.School of Mechanical Engineering,Northeastern Electric Power University,Jilin 132012,China)

**摘 要:** 针对现阶段油浸式变压器故障复杂性导致诊断精度不高的问题,提出一种基于主元分析(PCA)与麻雀搜索算法-轻量级梯度提升机(SSA-LightGBM)的油浸式变压器故障诊断方法。采集油中溶解气体数据,结合无编码比值方法构建17维故障特征矩阵,并对特征矩阵进行标准化处理得到联合特征。利用主元分析法进行特征融合,消除变量之间的信息冗余,构造融合特征。构建基于SSA-LightGBM变压器诊断模型,并采用十折交叉验证法验证该模型的分类能力。试验结果表明:提出的模型平均故障诊断精度为93.6%,与GA-LightGBM和GWO-LightGBM故障诊断模型相比,诊断精度分别提高了8.1和5.7百分点,验证了该方法能够有效提高油浸式变压器的故障诊断性能。

**关键词:** 故障诊断;变压器;主元分析;麻雀搜索算法;轻量级梯度提升机;麻雀搜索算法
**中图分类号:** TK 01:TM 41 **文献标志码:** A **文章编号:** 2097-0706(2023)03-0009-08

**Abstract:** In view of low diagnosis accuracy on oil-immersed transformer faults caused by the fault complexity at present, an oil-immersed transformer fault diagnosis method based on PCA and SSA-LightGBM is proposed. The data of dissolved gas in oil are collected, and a 17-dimensional fault feature matrix is constructed with codeless ratio method. The joint feature is obtained in the normalized matrix. Feature fusion is performed by using principal component analysis which can eliminate the redundant information between variables, and construct fusion features. The SSA-LightGBM transformer diagnosis model is constructed, and the ten-fold cross-validation method is used to verify the classification ability of the model. The experimental results show that the average fault diagnosis accuracy of the proposed model is 93.6%, which is 8.1 and 5.7 percentage points higher than that of GA-LightGBM and GWO-LightGBM fault diagnosis models, respectively, verifying that this method can effectively improve the fault diagnosis performance for transformers.

**Keywords:** fault diagnosis; transformer; principal component analysis; sparrow search algorithm; LightGBM; sparrow search algorithm

## 0 引言

油浸式变压器是电力系统中的重要设备,其运行状态决定电力系统能否安全可靠运行。由于变压器受到电应力、热应力、机械应力等多种因素的影响,变压器油和有机绝缘材料逐渐老化、分解,其物理性能等逐渐下降,产生碳氧化物和少量烷烃类气体。

因此,通过采集变压器油中溶解气体的监测数据,并结合适当方法对变压器故障类型做出精准诊断是可行的[1-5]。但是,由于变压器故障与油中气体质量分数之间缺乏明确的映射关系,提取变压器故障特征依然是目前油中溶解气体分析(Dissolved Gas Analysis,DGA)技术的主要难题。目前虽然形成了国际电工委员会(International Electrotechnical

**基金项目:** 吉林省科技发展计划项目(2021050959RQ)
Science and Technology Development Plan Project of Jilin Province(2021050959RQ)

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

Commission，IEC）三比值法、Rogers法及改良三比值法等方法，但上述方法的故障边界过于绝对且编码不全，导致变压器故障诊断结果的准确率不高，无法满足实际需求[6-7]。

随着人工智能研究的飞速发展，一些智能算法相继被引入变压器故障诊断中。目前，国内外学者利用DGA技术与支持向量机（Support Vector Machines，SVM）[8]、卷积神经网络[9]、模糊理论[10]等方法融合来提高故障诊断准确率，文献[11-13]采用多种智能算法优化SVM的惩罚因子$C$和核函数参数$g$能够提升其分类准确率，然而SVM是针对二分类问题提出的，在多分类方面存在不足之处。文献[14-16]采用卷积神经网络应用于变压器故障诊断，可有效诊断电力变压器的内部故障，但存在容易陷入局部极小点、收敛速度慢的问题。文献[17-20]利用模糊理论引入变压器故障诊断与状态评估中，解决故障原因、运行状态与故障机理之间的模糊性和不确定性问题，更好地提高变压器故障准确率，但由于凭其经验并借助于隶属函数进行处理，会存在“认知不确定”问题。

近年来，随着集成学习的不断发展，许多学者也将其应用至故障诊断中。轻量级梯度提升机（Light Gradient Boosting Machine，LightGBM）算法[21]则是集成学习中有效的分类算法之一，具有高效合理地处理类别型特征的特点，能在提升准确度的同时，提高运算速度并减少内存消耗。例如，文献[22]对原始DGA气体含量特征与比值特征相结合的故障相关特征进行筛选，将LightGBM模型进行参数优化，最终故障诊断准确率由88.02%上升至90.14%。文献[23]构造出16种气体体积比值作为LightGBM模型的输入故障特征，并使用网格搜索对LightGBM模型分别对学习率、树的深度、叶节点数和迭代次数寻优，最终分类准确率达到了93.39%。文献[24]采用LightGBM建立配电网10 kV馈线故障预测模型，经参数优化后气象、设备和运行因素故障精准度均超过90.00%。

基于以上研究，提出一种基于主元分析（Principal Component Analysis，PCA）[25]与麻雀搜索算法（Sparrow Search Algorithm，SSA）-LightGBM的油浸式变压器故障诊断方法，该方法对故障中产生的不同特征气体之间的关系进行挖掘，结合无编码比值法构建17维联合特征；利用Z-score法作归一化处理，利用主元分析法进行特征优选构造融合特征，消除变量信息之间的冗余特征；最后，结合十折交叉验证原理构建基于SSA-LightGBM的油浸式变压器故障诊断模型。

## 1 数据的预处理

### 1.1 预处理

油中溶解的气体主要包括$H_2$$C_2H_6$，$CH_4$，$C_2H_2$，$C_2H_4$。发生故障时通过特征气体之间的比例来判定变压器故障类型，根据DL/T 722—2014《变压器油中溶解气体分析和判断导则》可得出以下结论：（1）在局部放电中无$C_2H_2$，且$CH_4$较多；（2）油中电弧放电时，产生的气体以$H_2$和$C_2H_2$为主，伴有少量$CH_4$和$C_2H_4$；（3）中温过热时，变压器绝缘油中的总烃（TH=$CH_4$+$C_2H_2$+$C_2H_4$+$C_2H_6$）显著增多。

因此，采用无编码比值法构造17种特征参量[26-27]，其特征量见表1。其中，前5项分别为$H_2$，$CH_4$，$C_2H_4$，$C_2H_2$，$C_2H_6$的体积分数，其余为不同气体的体积分数比值。

表1 特征量及编号

Table 1 Characteristic quantities and their numbers

| 编号 | 特征量 | 编号 | 特征量 |
|---|---|---|---|
| 1 | $\varphi(H_2)$ | 10 | $\varphi(C_2H_6)/\varphi(TH)$ |
| 2 | $\varphi(CH_4)$ | 11 | $\varphi(H_2)/\varphi(TH)$ |
| 3 | $\varphi(C_2H_4)$ | 12 | $\varphi(CH_4)/\varphi(H_2)$ |
| 4 | $\varphi(C_2H_2)$ | 13 | $\varphi(C_2H_2)/\varphi(C_2H_6)$ |
| 5 | $\varphi(C_2H_6)$ | 14 | $\varphi(C_2H_4)/\varphi(C_2H_6)$ |
| 6 | $\varphi(TH)$ | 15 | $\varphi(C_2H_2)/\varphi(C_2H_4)$ |
| 7 | $\varphi(CH_4)/\varphi(TH)$ | 16 | $\varphi(H_2)/\varphi(H_2+TH))$ |
| 8 | $\varphi(C_2H_4)/\varphi(TH)$ | 17 | $\varphi(CH_4+C_2H_4)/\varphi(TH)$ |
| 9 | $\varphi(C_2H_2)/\varphi(TH)$ | | |

参考DL/T 722—2014《变压器油中溶解气体分析和判断导则》，数据标签分为正常状态、中低温热故障、中温热故障、高温热故障、高能放电、低能放电和局部放电。选取513例故障气体数据，构造17维联合特征矩阵，每行代表1个故障样本，如式（1）所示。

$$\boldsymbol{x}=\begin{bmatrix} x_{1,1} & \cdots & x_{1,17} \\ \vdots & \ddots & \vdots \\ x_{513,1} & \cdots & x_{513,17} \end{bmatrix}。\tag{1}$$

### 1.2 Z-score标准化处理

由于采集到的DGA数值分散，为不影响诊断的精度，需要将采集到的数据进行Z-score标准化，消除由于数据数量级相差较大造成的诊断误差。因此，对每种类型的特征进行标准化处理，

$$x_{i,j}^{*}=\frac{x_{i,j}-\mu_j}{\sigma_j},\ (i=1,\cdots,513;\ j=1,\cdots,17),\tag{2}$$

式中：$x_{i,j}^{*}$为标准化处理后联合特征；$\mu_j$为特征均值；$\sigma_j$为特征标准差；$x_{i,j}$为原始特征。经过标准化变换之后的每类故障特征平均数都为0，标准差都为1，特征集中程度和离散趋势一致，达到了消除特征之

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

间差的目的，即变压器故障样本特征经标准化处理后的联合特征为

$$\boldsymbol{x}^* = \begin{bmatrix} x^*_{1,1} & \cdots & x^*_{1,17} \\ \vdots & \ddots & \vdots \\ x^*_{513,1} & \cdots & x^*_{513,17} \end{bmatrix}。\tag{3}$$

## 2　基于PCA的融合特征矩阵构建

PCA法的核心思路是将高维数据映射到相互线性无关的新特征子空间，消除变量间的相关性，降低数据维度、简化复杂数据。

步骤1：求协方差矩阵$\boldsymbol{R}$。

$$\boldsymbol{R} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \ddots & \vdots \\ r_{n,1} & \cdots & r_{n,n} \end{bmatrix},\tag{4}$$

式中：$r_{n,n}$为相关系数。

步骤2：对协方差矩阵进行分解。

$$\boldsymbol{R} = \boldsymbol{USU}^{\mathrm{T}},\tag{5}$$

式中：$\boldsymbol{U}$为特征向量构成的矩阵；对角矩阵$\boldsymbol{S}$为奇异值矩阵。

步骤3：定义前$m$个主元的累积方差贡献率。

$$C = \frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i},\tag{6}$$

式中：主对角线中的元素$\lambda_i$是协方差矩阵$\boldsymbol{R}$的特征值，将特征值降序排列，即$\lambda_i$为各主元的方差。

图1为各主元的累积贡献率。由图1可见，前7个主元的累积贡献率包含了95%的主元信息。

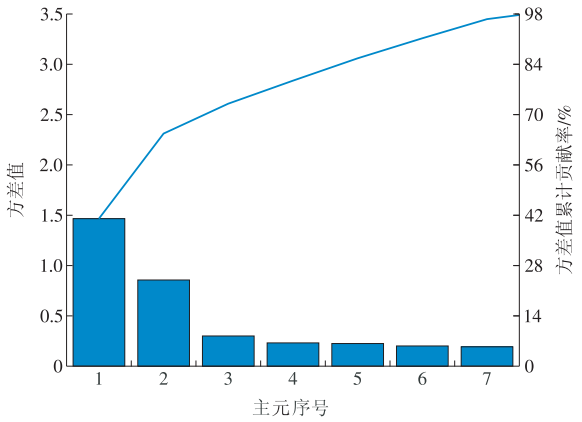



**图1　主元的累积贡献率**

**Fig. 1　Accumulation contribution rate of the main element**

因此，根据按累计贡献率确定主要主元的原则，选取前7个主元作为故障诊断的融合特征向量矩阵，记为

$$\boldsymbol{x}' = \begin{bmatrix} x'_{1,1} & \cdots & x'_{1,7} \\ \vdots & \ddots & \vdots \\ x'_{17,1} & \cdots & x'_{17,7} \end{bmatrix}。\tag{7}$$

对于采集新的样本，按特征向量构造方法，构造17维联合特征向量$\boldsymbol{x}_{\mathrm{new}}=[x_{\mathrm{new1}}, x_{\mathrm{new2}}, \cdots, x_{\mathrm{new17}}]$，然后向新的主元投影得到融合特征$\boldsymbol{X}_{\mathrm{new}}$，其投影算法为

$$\boldsymbol{X}_{\mathrm{new}} = [x_{\mathrm{new1}}, x_{\mathrm{new2}}, \cdots, x_{\mathrm{new17}}] \begin{bmatrix} x'_{1,1} & \cdots & x'_{1,7} \\ \vdots & \ddots & \vdots \\ x'_{17,1} & \cdots & x'_{17,7} \end{bmatrix}。\tag{8}$$

图2为验证融合特征向量$\boldsymbol{X}_{\mathrm{new}}$中不同主元分析数据的二维散点图。由图2可见，融合特征具有较好的聚类能力。

## 3　基于SSA-LightGBM构建故障诊断模型

### 3.1　LightGBM算法

LightGBM是由微软公司提出的基于决策树的Boosting算法，可应用于回归预测、故障监测、分类以及特征筛选等。

假设前一轮得到的学习器为$\boldsymbol{F}_{t-1}(x)$，损失函数为$L(y, \boldsymbol{F}_{t-1}(x))$，那么本轮的目标需要找到一个弱分类器$h_t(x)$，使得本轮的损失函数最小，即最小损失函数表示为

$$h_t(x) = \arg_{h \in H}\min\sum L\left(y, \boldsymbol{F}_{t-1}(x) + h_t(x)\right)。\tag{9}$$

然后拟合本轮损失函数的近似值，从而计算损失函数的负梯度，损失函数的近似值表示为

$$r_{ti} = -\frac{\partial L\left(y, \boldsymbol{F}_{t-1}(x_i)\right)}{\partial \boldsymbol{F}_{t-1}(x_i)}。\tag{10}$$

通过平方差近似拟合$h_t(x)$

$$h_t(x) = \arg_{h \in H}\min\sum\left(r_{ti} - h_t(x)\right)^2。\tag{11}$$

所得本轮的强学习器为

$$\boldsymbol{F}_t(x) = h_t(x) + \boldsymbol{F}_{t-1}(x)。\tag{12}$$

### 3.2　SSA算法

麻雀搜索算法是2020年提出的一种新型的群智能算法，不断更新个体位置从而模拟麻雀觅食和反捕食行为。

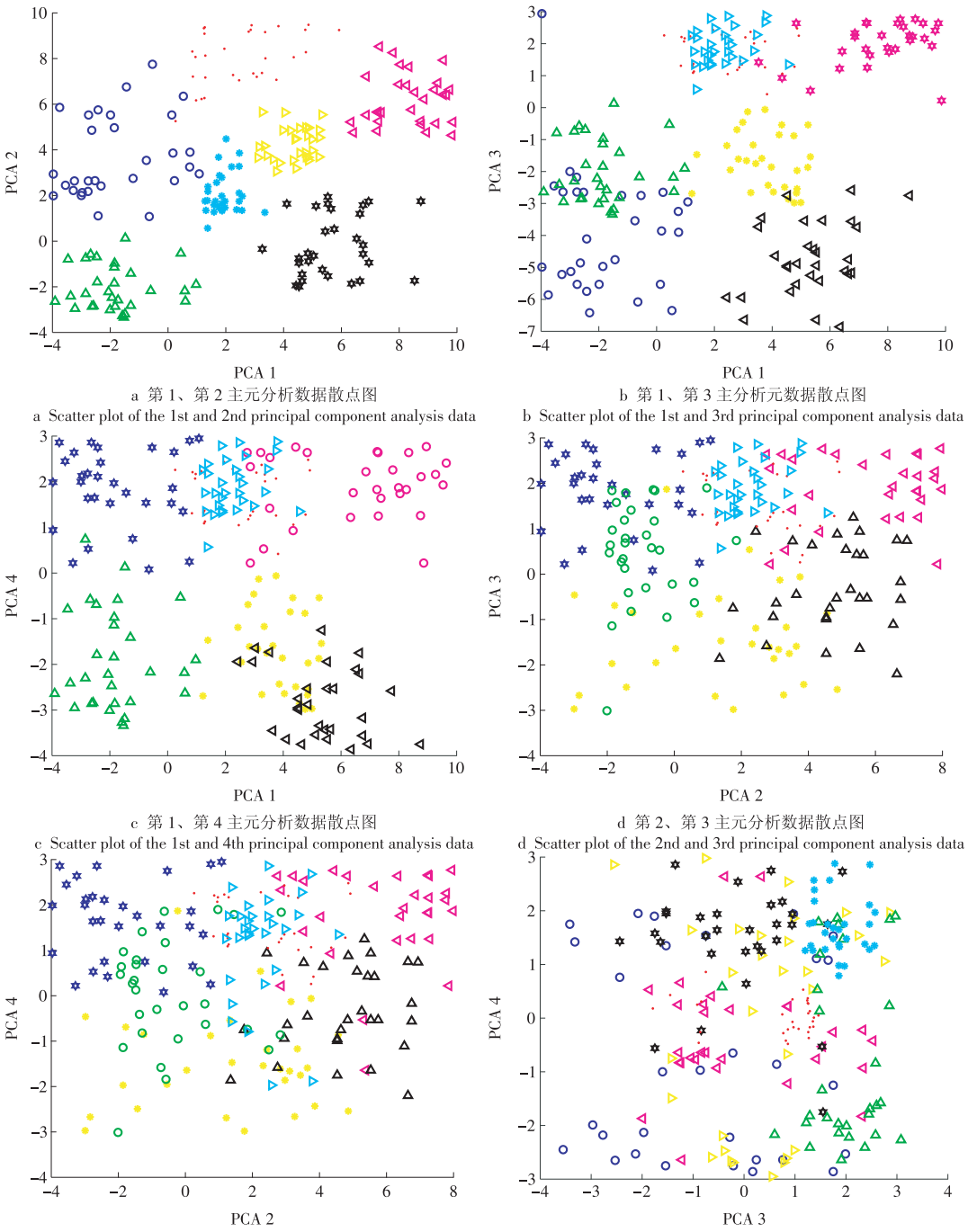
对麻雀种群与适应度值初始化：假设种群有$n$只麻雀，种群可表示为$\boldsymbol{X}=[x_1, x_2, \cdots, x_n]^{\mathrm{T}}$，麻雀各自对应的适应度函数为$\boldsymbol{F}=[f(x_1), f(x_2), \cdots, f(x_n)]$，具体表达式为

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nd} \end{bmatrix},\tag{13}$$

$$\boldsymbol{F}=\begin{bmatrix} f\left([x_{11}, x_{12}, \cdots, x_{1d}]\right) \\ f\left([x_{21}, x_{22}, \cdots, x_{2d}]\right) \\ \vdots \\ f\left([x_{n1}, x_{n2}, \cdots, x_{nd}]\right) \end{bmatrix}。\tag{14}$$

发现者占据整个麻雀种群的10%~20%，具有较好适应度值的发现者在搜索过程中会优先获取资

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.




a 第 1、第 2 主元分析数据散点图
a Scatter plot of the 1st and 2nd principal component analysis data

b 第 1、第 3 主分析元数据散点图
b Scatter plot of the 1st and 3rd principal component analysis data

c 第 1、第 4 主元分析数据散点图
c Scatter plot of the 1st and 4th principal component analysis data

d 第 2、第 3 主元分析数据散点图
d Scatter plot of the 2nd and 3rd principal component analysis data

e 第 2、第 4 主元分析数据散点图
e Scatter plot of the 2nd and 4th principal component analysis data

f 第 3、第 4 主元分析数据散点图
f Scatter plot of the 3rd and 4th principal component analysis data

**图 2　不同主元的二维散点图**

**Fig. 2　Two dimensions scatter of different principal component**

源，其位置更新为

$$x_{i,j}^{t+1}=\begin{cases}x_{i,j}^{t}\cdot\exp\left(\dfrac{-i}{\alpha\cdot N_{\text{iter, max}}}\right) & R_2<ST\\ x_{i,j}^{t}=Q\cdot\boldsymbol{L} & R_2\geqslant ST\end{cases},\quad(15)$$

式中：$t$ 为当前迭代次数；$x_{i,j}^{t+1}$ 为第 $i$ 只麻雀在第 $j$ 维的位置信息；$\alpha$ 为(0，1]的均匀随机数；$N_{\text{iter, max}}$ 为最大迭代次数；$R_2$ 为报警值，且 $R_2\in[0,1]$；$ST$ 为安全阈值，且 $ST\in[0.5,1.0]$；$Q$ 为服从正态分布的随机数；$\boldsymbol{L}$ 为 $1\times d$ 维的全 1 矩阵。当 $R_2<ST$ 时，表明捕食者未出现在觅食区域；当 $R_2\geqslant ST$ 时，表明捕食者出现并发出警报，所有发现者迅速飞往安全区域。

跟随者会在发现者的带领下进行局部搜索，具有较好适应度值的跟随者在搜索过程中会优先获取资源，其位置更新为

$$x_{i,j}^{t+1}=\begin{cases}Q\cdot\exp\left(\dfrac{x_{\text{worst}}^{t}-x_{i,j}^{t}}{i^2}\right) & i>\dfrac{n}{2}\\ x_p^{t}+\left|x_{i,j}^{t}-x_{i,j}^{t+1}\right|\boldsymbol{A}^{+}\boldsymbol{L} & i\leqslant\dfrac{n}{2}\end{cases},\quad(16)$$

式中：$x_p^{t+1}$ 为生产者当前最佳位置；$x_{\text{worst}}^{t}$ 为全局最差位

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

置；$\boldsymbol{A}$ 为 $1\times d$ 维矩阵，矩阵中的每个元素随机设定为1或-1，且 $\boldsymbol{A}^{+}=\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$。当 $i>n/2$ 时，表明第 $i$ 只加入者未能抢夺食物，需要前往其他区域觅食；而当 $i\leqslant n/2$ 时，表明加入者在最优个体 $x_{\mathrm{p}}$ 附近觅食。

更新警戒者的位置为

$$\begin{cases} x_{\mathrm{best}}^{t}+\beta\left|x_{i,j}^{t}-x_{\mathrm{best}}^{t}\right| & f_i\neq f_{\mathrm{g}} \\ x_{\mathrm{best}}^{t}+k\left(\dfrac{x_{i,j}^{t}-x_{\mathrm{best}}^{t}}{\left|f_i-f_{\mathrm{w}}\right|+\varepsilon}\right) & f_i=f_{\mathrm{g}} \end{cases}, \tag{17}$$

式中：$x_{\mathrm{best}}{}^{t}$ 为全局最优位置；$\beta$ 为步长控制函数，且服从均值为0，方差为1的正态分布函数；$k$ 为麻雀移动方向，且 $k\in[-1,1]$；$\varepsilon$ 为常数，避免分母为0；$f_i$ 为当前个体适应度值；$f_{\mathrm{g}}$ 和 $f_{\mathrm{w}}$ 分别为当前全局最优和最差的适应度值。

### 3.3　模型参数优化及训练测试

由于LightGBM模型在默认参数下具有诊断能力，但诊断准确率低，对LightGBM模型中特定超参数进行寻优可提升诊断准确率。其中，参数max_depth代表树的深度，合理调整可避免生成过深的树；feature_leaves代表特征的子抽样；learning rate代表学习率，其设定过小会导致梯度下降很慢，设定过大又会跨过最优值，产生振荡；subsample用于训练弱学习器，取值过小容易出现过拟合现象。主要参数寻优范围及最优值见表2。

**表2　LightGBM主要参数寻优范围及最优值**

**Table 2　Optimization range and optimal value of LightGBM key parameters**

| 参数名称 | 寻优范围 | 最优值 |
|---|---|---|
| max_depth | (4,12) | 6.000 |
| feature_leaves | (0,1) | 0.700 |
| learning rate | (0,1) | 0.814 |
| subsample | (0,1) | 0.632 |

采用SSA算法对LightGBM的参数进行优化，从而构建SSA-LightGBM诊断模型，其诊断流程如图3所示。

步骤1：对采集到的数据样本，利用无编码比值构建故障特征，并对其进行标准化处理，从而消除不同比值之间数值大小的影响。

步骤2：结合十折交叉验证法将样本数据分为训练集和测试集。

步骤3：设置LightGBM模型的初始参数及超参数优化范围，利用SSA算法对模型参数进行调整。

步骤4：计算麻雀新位置的适应度值，并更新整个麻雀种群所经历的最优和最劣适应度值及其位置。

步骤5：判断是否达到迭代次数，若是则终止迭代，存储当前得到的训练参数值作为最优参数值，否则返回步骤3。

步骤6：将存储的最优参数值输入LightGBM模型，利用测试集对模型的诊断效果进行测试，输出故障分类结果。

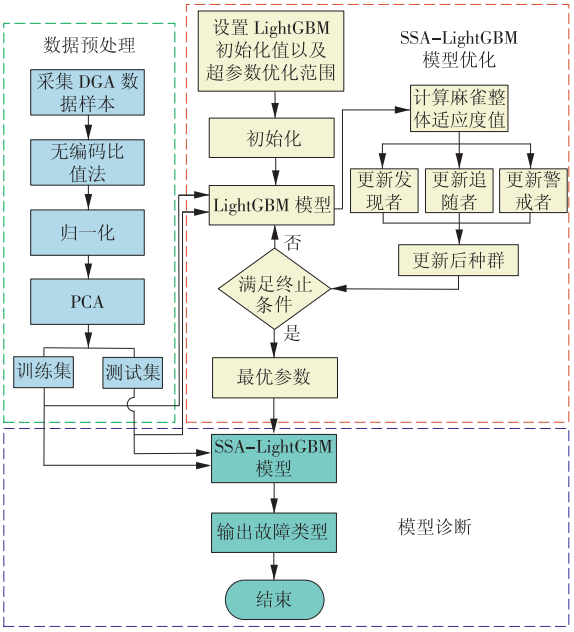



**图3　基于PCA与SSA-LightGBM变压器故障诊断流程**

**Fig. 3　Transformer fault diagnosis process based on PCA and SSA-LightGBMt**

### 3.4　诊断结果对比分析

#### 3.4.1　故障类型诊断结果分析

收集513例变压器故障时的特征气体，输入预设参数并结合十折交叉验证原理对所构建SSA-LightGBM诊断模型进行训练与测试，计算得到其平均故障诊断准确率为93.6%，总体故障诊断准确率为94.8%，见表3。

**表3　样本诊断结果及样本分布**

**Table 3　Sample diagnostic results and sample composition**

| 工作状态 | 样本总数 | 测试样本数 | 诊断正确数量 | 准确率/% |
|---|---|---|---|---|
| 正常状态 | 43 | 13 | 11 | 84.6 |
| 高温热故障 | 45 | 14 | 11 | 78.6 |
| 中温热故障 | 50 | 15 | 15 | 100.0 |
| 中低温热故障 | 128 | 39 | 38 | 97.4 |
| 低能放电 | 50 | 15 | 15 | 100.0 |
| 高能放电 | 78 | 23 | 23 | 100.0 |
| 局部放电 | 119 | 35 | 33 | 94.3 |
| 总体 | 513 | 154 | 146 | 94.8 |
| 平均 | — | — | — | 93.6 |

#### 3.4.2　多模型诊断结果分析

为验证LightGBM诊断模型的优越性，采用相同融合特征作为输入，将LightGBM模型诊断结果与k-近邻（k-Nearby Network，KNN）[28]，SVM[29]，后向传播（back propagation，BP）[30]模型诊断结果进行比较，

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

见表4。LightGBM模型平均准确率为90.3%，优于其他模型，验证了LightGBM模型的优越性，经SSA寻优后，准确率提高至93.6%，有效提高了诊断精度。

表4　不同模型诊断准确率

Table 4　Diagnostic accuracy of different models　%

| 模型名称 | KNN | SVM | BP | LightGBM | SSA-LightGBM |
|---|---|---|---|---|---|
| 诊断准确率 | 77.8 | 74.5 | 83.3 | 90.3 | 93.6 |

采用SSA对LightGBM诊断模型的相关超参数进行寻优，并与GA-LightGBM，GWO-LightGBM的优化结果进行对比分析，诊断结果见表5。GA-LightGBM，GWO-LightGBM和SSA-LightGBM的模型诊断准确率分别为85.6%，88.0%和93.6%，SSA-LightGBM模型的诊断精度最高，且中温热故障、低能放电和高能放电全部被正确识别，说明本文构建的SSA-LightGBM故障诊断模型能够有效提高变压器的故障诊断精度。

表5　多模型故障诊断结果

Table 5　Multi-model fault diagnosis results　%

| 工作状态 | 诊断准确率 | | |
|---|---|---|---|
| | GA-LightGBM | GWO-LightGBM | SSA-LightGBM |
| 正常状态 | 82.2 | 75.9 | 84.6 |
| 高温热故障 | 77.4 | 78.1 | 78.6 |
| 中温热故障 | 89.6 | 95.9 | 100.0 |
| 中低温热故障 | 87.1 | 92.1 | 97.4 |
| 低能放电 | 88.1 | 94.3 | 100.0 |
| 高能放电 | 87.6 | 90.4 | 100.0 |
| 局部放电 | 87.3 | 89.5 | 94.3 |
| 平均 | 85.6 | 88.0 | 93.6 |

### 3.4.3　不同特征量的诊断结果分析

为了验证经过PCA处理后优选出的7维融合特征的有效性，将试验采集的数据分别采用IEC三比值法、Rogers法和无编码比值法获得相应的故障诊断准确率，结果见表6，表明融合故障特征能够更深层次的挖掘了故障类型与DGA数据中的联系，可进一步提高故障诊断模型的准确率。

表6　不同特征量的平均准确率

Table 6　Average accuracy of different feature quantities　%

| 特征量名称 | IEC三比值 | Rogers | 联合特征 | 融合特征 |
|---|---|---|---|---|
| 平均准确率 | 87.4 | 86.9 | 88.4 | 93.6 |

## 4　结束语

针对变压器故障诊断精度较低的问题，建立了LightGBM的变压器故障诊断模型，其诊断结果与KNN，SVM和PNN模型相比具有更高的诊断精度。采用SSA优化LightGBM的相关超参数，能够进一步提高LightGBM变压器故障诊断模型的精度，并同GA-LightGBM和GWO-LightGBM相比，诊断精度分别提高了8.1和5.7百分点，能够更准确地对变压器故障信息做出精确、可靠的诊断，证明了本文所提变压器故障诊断方法的准确性与有效性。

采用无编码比值法构造17维故障特征，深层挖掘油中溶解气体与故障类型间的内在联系，利于变压器故障特征提取；PCA能够对故障特征降维，去除无效和冗余特征，提高特征提取后的诊断精度。

## 参考文献：

[1]韩志远，王广健，张晓静，等.基于故障模型的大型变压器故障综合诊断方法与应用[J].华电技术，2018，40(10)：22-26，77.
HAN Zhiyuan，WANG Guangjian，ZHANG Xiaojing，et al. Application of large transformer comprehensive fault diagnosis based on fault model [J]. Huadian Technology，2018，40(10)：22-26，77.

[2]马金平.变压器油中含气量超标检查与处理[J].华电技术，2018，40(7)：35-37，78.
MA Jinping. Inspection and processing of excessive gas content in transformer oil [J].Huadian Technology，2018，40(7)：35-37，78.

[3]康振华.大型电力变压器油色谱在线监测技术与案例分析[J].华电技术，2018，40(3)：17-20，77.
KANG Zhenhua. On-line monitoring technology and case analysis of large power transformer oil chromatography [J]. Huadian Technology，2018，40(3)：17-20，77.

[4]吴鑫杰.充氮气油浸式变压器取油样技术探讨[J].华电技术，2018，40(2)：46-47，79.
WU Xinjie. Discussion over oil sampling technique used in oil-immersed transformer filled with nitrogen [J]. Huadian Technology，2018，40(2)：46-47，79.

[5]龙思成，黄志鸿.基于多尺度极限融合网络的电力变压器故障诊断方法研究[J].综合智慧能源，2022，44(9)：78-83.
LONG Sicheng，HUANG Zhihong. Research on fault diagnosis method of power transformers based on multi-scale extreme fusion network [J]. Integrated Intelligent Energy，2022，44(9)：78-83.

[6]张镱议，彭鸿博，李昕，等.基于DGA特征量优选与改进磷虾群算法优化支持向量机的变压器故障诊断模型[J].电测与仪表，2019，56(21)：110-116.
ZHANG Yiyi，PENG Hongbo，LI Xin，et al. Fault diagnosis model of power transformers based on DGA features selection and improved krill herd algorithm optimization SVM [J].Electrical Measurement & Instrumentation，2019，56(21)：110-116.

[7]李黄曼，张勇，张瑶.基于ISSA优化SVM的变压器故障诊

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

断研究[J].电子测量与仪器学报，2021，35(3)：123-129.
LI Huangman, ZHANG Yong, ZHANG Yao. Study of transformer fault diagnosis based on improved sparrow search algorithm optimized support vector machine [J]. Journal of Electronic Measurement and Instrumentation, 2021，35(3)：123-129.

[8]余松，胡东，唐超，等.基于TLR-ADASYN平衡化数据集的MSSA-SVM变压器故障诊断[J].高电压技术，2021，47(11)：3845-3853.
YU Song, HU Dong, TANG Chao, et al. MSSA-SVM transformer fault diagnosis method based on TLR-ADASYN balanced data set [J]. High Voltage Engineering, 2021, 47(11)：3845-3853.

[9]朱瑞金，郝东光，胡石峰.小样本条件下基于卷积孪生网络的变压器故障诊断[J].电力系统及其自动化学报，2021，33(1)：64-69，84.
ZHU Ruijin, HAO Dongguang, HU Shifeng. Transformer fault diagnosis based on convolutional siamese network with small samples [J].Proceedings of the CSU-EPSA，2021，33(1)：64-69，84.

[10]禹建丽，周瑞芳.一种基于神经网络和模糊理论的变压器故障诊断[J].中南大学学报(自然科学版)，2013，44(S1)：243-247.
YU Jianli, ZHOU Ruifang. Transformer fault diagnosis based on neural network and fuzzy theory [J]. Journal of Central South University(Science and Technology)，2013，44(S1)：243-247.

[11]王晶，许素安，洪凯星，等.基于DGA特征量优选与GA-SVM的变压器故障诊断模型[J].变压器，2020，57(12)：36-40，46.
WANG Jing, XU Suan, HONG Kaixing, et al. Transformer based on DGA feature quantity optimization and GA-SVM fault diagnosis model[J].Transformer，2020，57(12)：36-40，46.

[12]方涛，钱晔，郭灿杰，等.基于天牛须搜索优化支持向量机的变压器故障诊断研究[J].电力系统保护与控制，2020，48(20)：90-96.
FANG Tao, QIAN Ye, GUO Canjie, et al. Research on transformer fault diagnosis based on a beetle antennae search optimized support vector machine [J].Power System Protection and Control，2020，48(20)：90-96.

[13]谢国民，倪乐水，曹媛.基于VSRP与β-GWO-SVM的变压器故障辨识方法[J].高电压技术，2021，47(10)：3635-3641.
XIE Guomin, NI Leshui, CAO Yuan. Transformer fault identification method based on VSRP and β-GWO-SVM [J].High Voltage Engineering，2021，47(10)：3635-3641.

[14]刘华，刘江永.基于图卷积神经网络的电力变压器故障诊断[J].湖南科技大学学报(自然科学版)，2021，36(3)：75-81.
LIU Hua, LIU Jiangyong. Fault diagnosis of power transformer based on graph convolutional network [J]. Journal of Hunan University of Science & Technology (Natural Science Edition)，2021，36(3)：75-81.

[15]杨德昌，廖文龙，任翔，等.基于胶囊网络的电力变压器故障诊断[J].高电压技术，2021，47(2)：415-424.
YANG Dechang, LIAO Wenlong, REN Xiang, et al. Fault diagnosis of transformer based on capsule network [J]. High Voltage Engineering，2021，47(2)：415-424.

[16]王涤，马爱军，归宇，等.基于P-CNN的局部放电绝缘故障融合诊断[J].高电压技术，2020，46(8)：2897-2905.
WANG Di, MA Aijun, GUI Yu, et al. Diagnosis of partial discharge insulation fault fusion based on P-CNN [J].High Voltage Engineering，2020，46(8)：2897-2905.

[17]吴瞻宇，董明，王健一，等.基于模糊关联规则挖掘的电力变压器故障诊断方法[J].高压电器，2019，55(8)：157-163.
WU Zhanyu, DONG Ming, WANG Jianyi, et al.Fault diagnosis of power transformer based on fuzzy association rules mining [J].High Voltage Apparatus，2019，55(8)：157-163.

[18]谢松，邹阳，蔡金锭.基于模糊粗糙集的变压器油纸绝缘状态评估[J].仪器仪表学报，2017，38(1)：190-197.
XIE Song, ZOU Yang, CAI Jinding.Assessment of transformer oil-paper insulation status with fuzzy rough set [J].Chinese Journal of Scientific Instrument，2017，38(1)：190-197.

[19]刘凯，彭维捷，杨学君.特征优化和模糊理论在变压器故障诊断中的应用[J].电力系统保护与控制，2016，44(15)：54-60.
LIU Kai, PENG Weijie, YANG Xuejun. Method of fault diagnosis for power transformer based on optimizing characteristics and the fuzzy theory [J]. Power System Protection and Control，2016，44(15)：54-60.

[20]王福忠，邵淑敏.基于模糊综合评价法的变压器运行状态评估[J].计算机仿真，2015，32(6)：141-145.
WANG Fuzhong, SHAO Shumin.Fuzzy strategy on running state evaluation of oil-immersed power transformer [J]. Computer Simulation，2015，32(6)：141-145.

[21]柴琴琴，陈舒迪，王武，等.基于改进GWO-LightGBM的滚动轴承未知故障诊断[J].航空动力学报，2022，37(4)：848-855.
CHAI Qinqin, CHEN Shudi, WANG Wu, et al. Unknown fault diagnosis of rolling bearings based on improved GWO-LightGBM [J]. Journal of Aerospace Power, 2022, 37(4)：848-855.

[22]王凯.基于集成学习的变压器故障预测方法研究[D].济南：山东大学，2021.
WANG Kai. Research on transformer fault prediction method based on ensemble learning [D]. Jinan: Shandong University，2021.

[23]袁淼.基于油中气体分析的变压器故障智能诊断方法研究[D].广州：广东工业大学，2021.
YUAN Miao. Research on intelligent diagnosis method of

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.

transformer fault based on dissolved gas analysis [D]. Guangzhou:Guangdong University of Technology,2021.

[24]刘亚玲.基于深度学习的10 kV馈线故障预测方法研究[D].重庆:重庆大学,2019.
LIU Yaling. Research on 10 kV feeder fault prediction method based on deep learning [D].Chongqing:Chongqing University,2019.

[25]覃煜,黄慧红,方健,等.基于FCM和改进PCA的变压器故障诊断[J].高压电器,2018,54(12):262-267.
QIN Yu,HUANG Huihong,FANG Jian,et al.Fault diagnosis of transformer based on FCM and improved PCA [J].High Voltage Apparatus,2018,54(12):262-267.

[26]张育杰,冯健,李典阳,等.基于油色谱数据的变压器故障征兆新优选策略[J].电网技术,2021,45(8):3324-3331.
ZHANG Yujie, FENG Jian, LI Dianyang et al. New feature selection method for transformer fault diagnosis based on DGA data [J]. Power System Technology,2021,45(8):3324-3331.

[27]白星振,臧元,葛磊蛟,等.变压器故障诊断用油中溶解气体征兆优选方法[J/OL].高电压技术:1-16(2022-09-08)[2022-10-04].https://doi.org/10.13336/j.1003-6520.hve.20220670.
BAI Xingzhen,ZANG Yuan,GE Leijiao,et al. Selection method of feature derived from dissolved gas in oil for fault diagnosis of transformers [J/OL].High Voltage Engineering:1-16(2022-09-08)[2022-10-04].https://doi.org/10.13336/j.1003-6520.hve.20220670.

[28]张彼德,梅婷,王涛.电力变压器故障诊断的k值自适应加权KNN算法研究[J].湖北电力,2020,44(2):6-12.
ZHANG Bide,MEI Ting,WANG Tao.Research on adaptive k-value weighted KNN algorithm for power transformer fault diagnosis [J].Hubei Electric Power,2020,44(2):6-12.

[29]刘裕舸,黄忠辉.基于SVM算法优化的变压器故障诊断研究[J].红水河,2022,41(2):89-93.
LIU Yuge, HUANG Zhonghui.Transformer fault diagnosis based on SVM algorithm optimization [J].HongShui River,2022,41(2):89-93.

[30]王妍,张太华.基于数字孪生的优化概率神经网络变压器故障诊断[J].组合机床与自动化加工技术,2020(11):20-23.
WANG Yan,ZHANG Taihua.Fault diagnosis of transformers based on optimal probabilistic neutral network based on digital twin [J]. Modular Machine Tool & Automatic Manufacturing Technique,2020(11):20-23.

（本文责编:惠忻）

**收稿日期**:2022-06-10;**修回日期**:2022-10-10
**上网日期**:2023-03-25;**附录网址**:www.iienergy.cn

**作者简介**:

肖宏磊(1985),男,高级经济师,从事电力系统工程和自动化技术研究,18244891@qq.com;

杨海琦*(1997),女,在读硕士研究生,从事输配电设备智能化、机电设备故障诊断的研究,xhaiqi0526@163.com。

*为通信作者。

© Editorial Department of Integrated Intelligent Energy. This is an open access article under the CC BY-NC-ND license.