

# 基于PCA的火电厂电气设备故障诊断方法

聊方伦<sup>1</sup>, 周平<sup>2</sup>

(1. 安徽安庆皖江发电有限公司, 安徽 安庆 246008;

2. 浙江浙能嘉华发电有限公司, 浙江 平湖 314201)

**摘要:** 电气设备故障诊断中对样本数据要求较高,且易出现特征提取难、诊断精准度低等问题,因此,提出基于PCA的火电厂电气设备故障诊断方法。通过PCA分析得出数据间相关性,经过标准化计算确定主元数量,利用最优重构法构建主元模型,基于PCA处理设备输出相应的采样信号,提取故障特征主要成分,采用BP神经网络与遗传算法相结合,优化原网络中参数结构,获取出最优取值网络结构、初始权值、阈值以及学习速率,构建出具有最佳结构和参数的神经网络实现故障诊断。仿真实验结果表明,所提方法可通过最优网络结构减少整体计算步骤,测得实验数据与真实值相差较小,具有较高有效性。

**关键词:** PCA算法; 电气设备; 故障诊断; BP神经网络; 遗传算法

中图分类号: TP183 文献标识码: A 文章编号: 1003-7241(2022)03-0039-05

## Fault Diagnosis Method of Electrical Equipment in Thermal Power Plant Based on PCA

LIAO Fang-lun<sup>1</sup>, ZHOU Ping<sup>2</sup>

(1. Anhui Anqing Wojoint Power Generation Co., Ltd., Anqing 246008 China;

2. Zhejiang Zheneng Jiahua Power Generation Co., Ltd., Pinghu 314201 China )

**Abstract:** For the electrical equipment fault diagnosis methods, it is easy to have high sample data requirements, difficult feature extraction and low diagnostic accuracy. Therefore, this paper proposes a PCA based fault diagnosis method for electrical equipment in thermal power plant. The correlation between the data is obtained by PCA method, and the number of principal components is determined by standardized calculation. The principal component model is constructed by using the optimal reconstruction method. The sampling signal based on PCA processing equipment output response is processed by PCA, and the main components of fault characteristics are extracted. BP neural network and genetic algorithm are combined to optimize the parameter structure of the original network and obtain the optimal value network Structure, initial weight, threshold and learning rate, the neural network with the best structure and parameters is constructed to realize fault diagnosis. The simulation results show that the proposed method can reduce the overall calculation steps through the optimal network structure, and the difference between the experimental data and the real value is small, so it has high effectiveness.

**Key words:** PCA algorithm; electrical equipment; fault diagnosis; BP neural network; genetic algorithm

### 1 引言

我国工业技术快速发展,相对应的生产装置也呈现出大型、高效以及自动化趋势发展。在现实应用中相关用户对设备要求也在不断提升,在保证设备高效节能的同时,还需要保证设备在正常运行中减少故障产生,但由于大多数电气设备都是高科技产品,在实际运用中难以

避免出现故障问题。诊断技术一般是在设备正常运行状态下,寻找出运行出现异常的地方,并根据故障数据判定相对应的解决对策,以此完成诊断。诊断的目的是在保证人员安全的同时,尽量减少因故障产生的影响,也是为了保障设备的基本安全,延长其使用寿命,从根本上提升设备生产效率,获取出最大化的经济效益。

黄葆华等人利用SBM和FPSDG的优点,将两者融合

得到基于SBM-FPSDG的设备故障监测诊断方法,以高压加热器为研究对象,利用其历史运行数据建立SBM模型,根据设备运行特性建立了故障单元库和概率,可以实现高压加热器监测及故障诊断<sup>[1]</sup>。周璐婕等人提出一种基于反向传播BP神经网络的车载设备智能故障诊断方法。为了避免主特征信息的冗余,采用主成分启发式属性约简算法进行降维降噪。另外,由于BP神经网络对初始权值比较敏感,采用遗传算法对BP神经网络的初始权值和阈值进行了优化<sup>[2]</sup>。

虽然上述两种方法最后可以实现故障诊断,但由于该方法并不会针对故障特征提取分析,所以会出现计算量大的问题,因此本文提出基于PCA的火电厂电气设备故障诊断方法。PCA(principal component analysis,主元分析)作为一种不依赖于系统数学模型的建模方法,它可以将多个相关变量同时转化为几个不相关的变量,为火电厂电气设备故障诊断提供一种新的途径。引用BP神经网络与遗传算法相结合的诊断方法,利用其参数及结构完成诊断。

## 2 PCA下火电厂电气设备故障诊断

### 2.1 PCA主元数量确定

该方法可以根据不同的正交向量,让计算过程中的数据变化呈现在向量方向中,以便获取出传感器数据之间的相关性<sup>[3]</sup>。

首先在电气设备正常运行时,选取出一段传感器采集数据,并把数据构成数据矩阵 $\bar{X}_{n \times m}$ ,在矩阵集合中将 $n$ 和 $m$ 描述为采样点数以及传感器数量。为了能够有效除掉因量纲不同带来的不合理影响,便需要对 $\bar{X}_{n \times m}$ 进行标准化计算,获取出对应数据集合 $\bar{X}_{n \times m}$ ,构建PCA的过程统计模型如下所示:

$$X_{n \times m} = \hat{X}_{n \times m} + \tilde{X}_{n \times m} = \hat{T}_{n \times l} \hat{P}_{m \times l}^T + \tilde{T}_{n \times (m-l)} \tilde{P}_{m \times (m-l)}^T \quad (1)$$

式中: $\hat{X}_{n \times m}$ 为 $X_{n \times m}$ 的估计; $\tilde{X}_{n \times m}$ 为残差矩阵; $\hat{T}_{n \times l}$ 和 $\tilde{T}_{n \times (m-l)}$ 分别为主元得分和负荷矩阵; $\hat{P}_{m \times l}^T$ 和 $\tilde{P}_{m \times (m-l)}^T$ 分别为残差得分和负荷矩阵; $l(l < m)$ 为主元个数。

$X_{n \times m}$ 的分解方式是确保 $P_{m \times m} = [\hat{P}_{m \times l}, \tilde{P}_{m \times (m-l)}]$ 是正交阵,通过对矩阵 $X_{n \times m}$ 的协方差矩阵 $R = X_{n \times m}^T X_{n \times m} / (n-1)$ 进行特征向量分析,得到矩阵 $P_{m \times m}$ ,则有:

$$\begin{aligned} R &= P_{m \times m} \sum P_{m \times m}^T T_{n \times m} = [\hat{T}_{n \times l}, \tilde{T}_{n \times (m-l)}] \\ &= X_{n \times m} P_{m \times m} = [X_{n \times m} \hat{P}_{m \times l}, X_{n \times m} \tilde{P}_{m \times (m-l)}] \end{aligned} \quad (2)$$

式中,对角矩阵 $\Sigma = \text{diag}(\lambda_i)$ , $\lambda_i(i=1,2,\dots,l)$ 为 $X_{n \times m}$ 在新坐标系 $P_{m \times m}$ 方向上的方差。

根据上式计算结果将式(2)代入式(1)便有:

$$X_{n \times m} = X_{n \times m} \hat{P}_{m \times l} \hat{P}_{m \times l}^T + X_{n \times m} \tilde{P}_{m \times (m-l)} \tilde{P}_{m \times (m-l)}^T \quad (3)$$

$$= X_{n \times m} \hat{C}_{m \times m} + X_{n \times m} \tilde{C}_{m \times m}$$

式中 $\hat{C} = \hat{P} \hat{P}^T$ , $\tilde{C} = \tilde{P} \tilde{P}^T$ 。

计算结果的主元数量精准程度,可以直接影响到后续步骤的实际取值结果,所以创建系统主元模型中,需要以确定主元数量为基本前提,此处本文采用最优重构法<sup>[4]</sup>,通过下列计算确认构建其模型的主元数量。该方法定义不可重构方差为:

$$u_j = \frac{\zeta_j^T (I - \hat{C}) \sum (I - \hat{C}) \zeta_j}{[\zeta_j^T (I - \hat{C}) \zeta_j]^2}, i=1,2,\dots,m \quad (4)$$

在上式中将 $\zeta_j$ 描述为电气设备故障矩阵 $\zeta \in I_{m \times m}$ 的第 $j$ 列向量。

根据上式计算结果可以明显看出, $u_i$ 的取值结果与 $\hat{C}$ 有明显的关系,而 $\hat{C}$ 则和计算主元个数有着直接联系。为了可以在每个故障数据方向上都能实现重构计算,本文根据下式计算结果,挖掘出可以符合其所有条件的主元个数来判定是否为最优重构,有:

$$l = \min \left( \sum_{j=1}^a u_j \right) \quad (5)$$

式中将 $a$ 和 $l$ 分别描述为计算中可能发生变化数据的数量以及实际存在的主元数量。通过计算不同的主元个数,即可分别获取出 $\sum_{j=1}^a u_j$ ,这样即可得知选取与其结果相对应的主元数量为最优主元数。

根据上式计算结果以及主元数据的所有投影,把最开始 $m$ 个数据变量降低到目前这些 $l$ 个主元, $l$ 维的主元空间便替换的之间的 $m$ 维过程数据空间,还会除掉两个空间之间的相关性<sup>[5]</sup>。

### 2.2 故障特征提取

结合上述计算分析,采用PCA对故障特征数据提取降维计算过程如下:

(1) 原始数据标准化:

假设令初始数据采集样本表示为 $X = (x_{ij})_{a \times b}$ ,那么为了能够有效解决存在数据中的变量,因数值结果差异大、量纲不同等问题带来的不利影响,在计算中需要将其数据进行标准化处理,便有:

$$y_{ij} = \frac{x_{ij} - x_j}{S_j}, i=1,2,\dots,m; j=1,2,\dots,n \quad (6)$$

$$\text{式中 } x_j = \frac{1}{m} \sum_{i=1}^m x_{ij}, S_j = \sqrt{\frac{\sum_{i=1}^m (x_{ij} - x_j)^2}{m-1}}$$

根据上式计算结果,即可获取出新的标准化矩阵:

$$Y = (y_{ij}) \quad (7)$$

(2) 构建相关矩阵,求解特征取值与向量

根据上述获取出的样本相关系数矩阵  $R = Y_{m-1}^T Y$ , 计算

求解出其特征值  $\lambda_i$  以及向量  $\sigma_i (i=1,2,\dots,m)$ 。

根据上述计算可得知,PCA 不仅可以减少后续计算的次数,而且可以提高整体性能<sup>[6]</sup>。

### 2.3 基于BP神经网络和遗传算法的故障诊断

经过上述计算得知,BP神经网络在电气设备故障诊断过程中,极为容易出现误差,导致后续诊断失误情况,因此本文提出遗传算法与BP神经网络相结合。其基本运算过程如图1所示。

遗传算法以及BP神经网络都是基于生物学原理研究的数学计算方法,本文在诊断故障设备的过程将其结合,是为了能够从中获取出两种算法不同优点,在有效诊断的同时,得出有效的解决途径<sup>[7]</sup>。

在实际计算过程中,首先利用遗传算法对BP网络中的结构及参数进行优化处理,从而获取出最优取值的网络结构、初始权值、阈值以及学习速率。

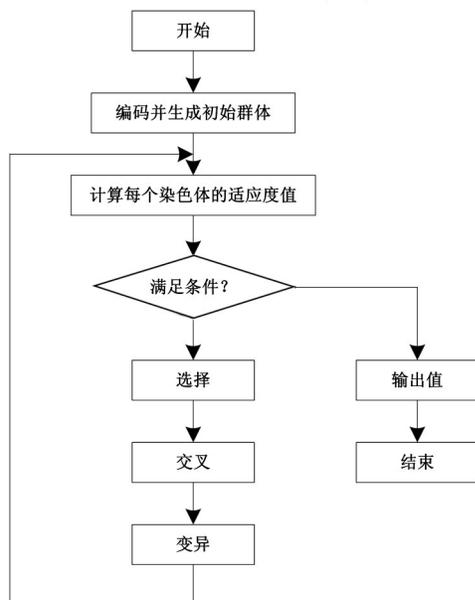


图1 遗传算法流程图

#### 2.3.1 结构与算法过程

计算过程中可以将BP算法分为正向传播、反向传

播,其工作过程如下:在正向传播中,样本从输入层通过所有隐藏层传输到输出层,在输出层比较当前输出和预期输出。如果满足误差要求,则完成计算;如果不满足,则进行反向传播<sup>[8]</sup>。

在正向传播计算过程中,首先需要将数据样本从输入层转换到输出层,转到输出层之后将目前计算结果和实际期望取值进行对比,经对比结果得知,如果符合误差条件要求,那就结束计算,如果不满足,则执行反向传播;再现将返回一个错误信号,并修改单个神经元阈值以将误差最小化<sup>[9]</sup>。

其中具体计算流程如下述:

第一步:选取出合适的输入样本  $X_k$  和目标样本  $T_k$ , 并将其代入至神经网络计算中;

第二步:正向计算。正向计算过程中分别将输入层第  $i$  个神经元输出以及第  $h$  个神经元输出表示为:

$$Y_i = f(x_i) \quad (8)$$

$$I_h = \sum_n \omega_{hi} I_i + \theta_i \quad (9)$$

第三步:计算输出层第  $i$  个神经的输出以及输入

$$I_j = \sum_h \omega_{jh} Y_h + \theta_j \quad (10)$$

$$Y_j = f(I_j) \quad (11)$$

第四步:求解输出误差

$$E_k = \frac{1}{2} \sum_{j=1}^M (d_j - Y_j)^2 \quad (12)$$

第五步改善输出层、隐含层的全部阈值、权值:

$$\omega'_{jh} = \omega_{jh} + \Delta\omega_{jh} = \omega_{jh} + \eta \delta_j^M Y_h$$

$$\omega'_{hi} = \omega_{hi} + \Delta\omega_{hi} = \omega_{hi} + \eta \delta_h^H Y_i$$

$$\theta'_h = \theta_h + \Delta\theta_h = \theta_h + \eta \delta_j^M$$

$$\theta'_j = \theta_j + \Delta\theta_j = \theta_j + \eta \delta_h^H \quad (13)$$

式中分别将  $\delta_j^M = Y_j(1-Y_j)(d_j - Y_j)$  和  $\delta_h^H = f'(I_h) \sum_{j=1}^M$

$\delta_j^M \omega_{jh}$  描述为输出层神经元的部分梯度以及隐含层神经元的部分梯度<sup>[10]</sup>。

第六步:计算BP神经网络总误差

$$E = \sum_{k=1}^K E_k = \frac{1}{2} \sum_{k=1}^K \sum_{j=1}^M (T_j^k - Y_j^k)^2 \quad (14)$$

第七步:判断是否达到预定的培训时间,如果符合条件则结束训练,反之,返回步骤二继续训练。

#### 2.3.2 隐层节点数和学习速度确定

由于神经网络的隐层节点数和学习速度是由实验测试确定的,所以在确定隐层节点数时,学习速度为0.1。对5-10个隐层节点的BP网络进行了10次全训练,分别计算了平均运行长度和训练转化率。表1列出了不同隐层节点的BP网络的平均转换阶段数和平均工作时间,即10个培训班的培训专业数<sup>[11-12]</sup>。

表1 不同隐含层节点数的平均收敛次数及运行时间

节点数量	5	6	7	8	9	10
运行时间(s)	1.98	1.85	1.78	1.60	1.65	1.74
收敛数量	58	49	43	36	38	43

由表1可知,当隐层节点数为8时,网络收敛速度较低,工作时间较短,无故障训练。因此,神经网络层BP的隐节点数被定义为8个。

表2 不同学习速率的平均收敛次数及运行时间

学习速率	0.03	0.06	0.1	0.2	0.3	0.4
运行时间(s)	1.87	1.85	1.65	1.62	1.66	1.74
收敛步数	44	46	42	39	41	51

由表2可知,当网络中的学习率为0.1-0.3时,网络中的收敛阶段数较少,工作时间也较短,且无训练失败。在此之后,BP神经网络的学习速度为0.15。

### 2.3.3 故障诊断输出

其中具体计算流程如下:

步骤一:遗传算法中的每个种群集合都将代表BP网络中的一个结构及参数;

步骤二:保持在数据初始状态下,构建BP神经网络;

步骤三:BP网络输入层中输入一组遗传数据样本,并利用遗传样本训练,获取出神经网络的输出误差取值;

步骤四:根据上一步进行检测,查看是否有遗漏没有计算的样本,如果没有则将步骤三的取值,将其看作为适应度函数,然后分别计算每个样本的适应度取值;

步骤五:针对遗传算法中的数据交叉可能性进行计算,利用交叉以及突变的遗传方法产生新的种群;

步骤六:根据步骤五的结果,若是不符合下一步骤的条件则转为上述步骤三,符合条件下一步;

步骤七:结束遗传基本,计算出目前每个数据集合的适应函数取值,并在众多取值中选取最高值,看为最优单位。根据数据单元解码计算,获取出最优神经网络结构、初始权值、阈值以及学习速率,并根据其结果构建最

优神经网络。

步骤八:最优BP网络中输入层采样一组数据样本,计算得出训练网络的实际输出;

步骤九:再次检测数据是否经过训练计算,如果没有返回步骤八;

步骤十:计算误差与实际是否差距过大,若是满足 $E < \sigma$ 条件就可以结束训练计算<sup>[13-14]</sup>。

在实际应用中本文根据编码方式随机产生含有80个个体的初始群体,每个个体的编码长度为 $25 + ((8 \times 25 + 25 \times 4 + 25 + 4) + 1)$ 。其中,前面的是25个是连接基因,采用二进制编码方式,每个个体初始化为0或1;后面的 $((8 \times 25 + 25 \times 4 + 25 + 4) + 1)$ 个是参数基因,采用实数编码方式。其中, $(8 \times 25 + 25 \times 4 + 25 + 4)$ 个是权阈基因,初始化为 $(-1, 1)$ 的实数,最后的一个是速率基因,初始化为 $(0, 1)$ 的实数。

## 3 仿真实验

为了能够进一步验证本文方法的有效性,将在实际仿真环境中,对电气设备中的轴承进行故障诊断验证。一般情况下大型电气设备轴承是比较容易出现故障的地方,所以在实际仿真中采用对轴承对的检测来进一步验证本文方法的实际有效性,以及检测故障的诊断能力。

实际仿真数据采用某数据库提供的轴承故障数据集,然后将其分为三种,分别是外圈故障、内圈故障以及滚珠故障。从三个数据集中分别抽取1500个样本,其中1000个样本进行训练,另外500个用于测试。

诊断过程中分别为将本文方法与文献方法[1]、[2]进行诊断精准度对比,其中对比结果如下表所示。

表3 故障诊断准确度对比表/%

项目	文献[1]方法		文献[2]方法		本文方法	
	A	B	A	B	A	B
外圈故障	88.21	89.37	89.53	90.44	95.97	97.27
内圈故障	91.89	92.26	92.11	94.33	94.17	95.43
滚珠故障	85.92	87.91	91.77	92.76	96.22	97.82
平均准确率	88.67	89.85	89.84	92.51	95.45	96.84

注:A为训练样本;B测试样本

从上表中即可看出,与文献方法相比,本文方法在三种故障方面都具有较高的诊断率,由此可得知本文方法所使用的PCA进行故障特征提取时,具有较大的优势。

将网络诊断实际输出和期望输出进行对比,输出结

果如图2所示。其中神经网络隐含层节点数、收敛步数、学习速率以及网络误差分别设置为13、9、0.182以及0.048。

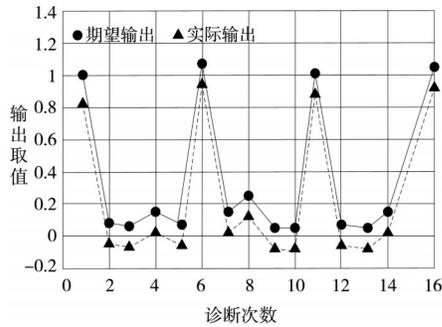


图2 故障诊断输出对比图

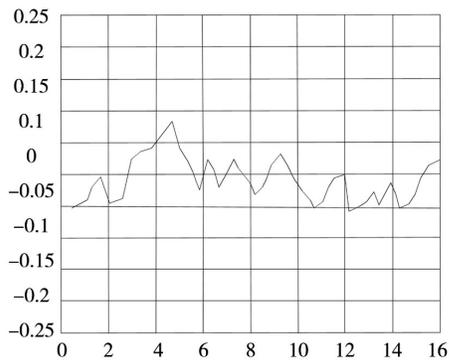


图3 故障诊断误差曲线图

从上述图2可以看出,本文算法输出结果与实际值相差极小,同时图3能够看出,本文算法经过多次计算,误差都能有效保持在0.1之下,获得的结果误差在实际运用过程中可忽略不计,可广泛使用。

## 4 结束语

由于电气设备故障诊断方法,容易出现因故障特征提取困难导致诊断失误等情况,本文提出基于PCA的火电厂电气设备故障诊断方法。根据PCA算法计算得知相关性,确定主元数量,并构建主元分析模型,随后采用BP神经网络进行故障诊断,计算中发现BP用于诊断会出现隐含层节点数难判定的情况,基于此本文采用GA+BP相结合算法,构建合适的网络结构进行诊断。虽然可以有效解决其缺陷,但在故障提取计算中,需要严谨的计算,因为主元数量一旦有差异,那么后续步骤都会失误,所以仍有待优化。

## 参考文献:

[1] 黄葆华,仇晓智,周卫庆,等.基于相似性建模和模糊概率有向图的热力设备故障诊断[J].热力发电,2018,47(4):108-113.

[2] 周璐婕,董昱.基于GA-BP神经网络的列控车载设备故障诊断方法研究[J].铁道科学与工程学报,2018,15(12):243-251.

[3] 王靖宇,吴迪,周玉权.机械程序锁在电气设备上的安装和使用[J].水电站机电技术,2019,42(9):51-54.

[4] 黄鑫,陈仁祥,黄钰.卷积神经网络在机械设备故障诊断领域应用与挑战[J].制造技术与机床,2019(1):96-100.

[5] 上官伟,孟月月,杨嘉明,等.基于LSTM-BP级联网络的列控车载设备故障诊断[J].北京交通大学学报,2019,43(1):59-67.

[6] 唐立力,陈国彬.基于MEA优化BP神经网络的农机滚动轴承故障诊断[J].农机化研究,2019,41(3):214-218.

[7] 李正飞,谭泽汉,陈焕新,等.基于ReliefF与mRMR耦合特征选择的多联机制冷剂充注量故障诊断[J].暖通空调,2018,48(10):53-58.

[8] 权凌霄,郭海鑫,盛世伟,等.采用“GA+LM”优化BP神经网络的电液伺服阀故障诊断[J].中国机械工程,2018,29(5):505-510.

[9] 郭梦茹,谭泽汉,陈焕新,等.基于遗传算法和BP神经网络的多联阀门类故障诊断[J].制冷学报,2018,39(2):119-125.

[10] 何春,李琦,吴让好,等.基于故障传播的模块化BP神经网络电路故障诊断[J].计算机应用,2018,38(2):602-609.

[11] 雷亚国,贾峰,孔德同,等.大数据下机械智能故障诊断的机遇与挑战[J].机械工程学报,2018,54(5):94-104.

[12] 姜媛媛,张书婷.基于改进的VMD和CNN神经网络的光伏逆变器软故障诊断方法研究[J].电测与仪表,2021,58(2):158-163.

[13] 白瑞飞,张伟,赵豆豆.基于IPSO-FNN的风电齿轮箱故障诊断方法[J].传感器与微系统,2021,40(1):125-128.

[14] 张绍辉,罗洁思.基于频谱包络曲线的稀疏自编码算法及在齿轮箱故障诊断的应用[J].振动与冲击,2018,37(4):249-256.

作者简介:聊方伦(1976-),男,本科,高级工程师,从事发电厂生产管理工。