



· 燃气输配与储运 ·

基于支持向量机集合的油浸式变压器故障诊断

肖海燕¹, 孙爱伦²

(1. 中国市政工程华北设计研究总院有限公司 第十设计研究院, 天津 300074;

2. 山东海成石化工程设计有限公司天津分公司, 天津 300384)

摘要: 油浸式变压器故障诊断是多分类问题。为了解决传统油浸式变压器故障诊断需要样本数据量大的问题,提出基于支持向量机集合的变压器故障诊断方法。将多分类问题分解为多个二分类问题的组合,变压器故障分为5个类别,构建10个支持向量机,每个支持向量机对应2个故障类别。样本每经过1个支持向量机都会获得1个可能的所属类别,当经过所有支持向量机后,通过投票法确定样本最终所属类别。通过模型训练,求得核函数的取值,建立模型。模型测试结果正确率达到86.25%,验证了方法的有效性,说明该方法适用于油浸式变压器故障诊断。

关键词: 油浸变压器; 故障诊断; 油中溶解气体分析; 支持向量机集合

中图分类号: TU996.6 **文献标志码:** A **文章编号:** 1000-4416(2021)06-0B01-03

1 概述

水电厂变压器作为电网的核心设备,其平稳运行直接关系到整个电力系统的安全稳定,因此准确地评价变压器的运行状态,准确预测并提前定位可能的故障点至关重要。目前主要的变压器故障诊断方法都是基于油中溶解气体分析方法(Dissolved Gas Analysis, DGA),该方法既可以通过随时抽取油样进行气体组分分析,又可以通过色谱分析对变压器的运行状态进行实时监控。伴随着信息技术的发展,各种新的技术与DGA相结合的故障诊断方法逐步出现,其中最主要的是基于神经网络的诊断方法。

Miranda等人^[1]首次将神经网络算法引入到变压器故障诊断过程,首先建立了3层的神经网络,将油中溶解气体的含量离散化,作为输入变量,将故障类型作为输出变量,通过训练样本的学习得到权重系数和阈值,最终将模型运用到实际数据,效果良好。针对神经网络算法收敛速度慢的特点,杨道武等人^[2]引入了粒子群算法加快算法的收敛速度,同时不影响原有算法的搜索能力,提高了神经网络算

法在实际条件下的应用效果。

从以上研究过程来看,神经网络在油浸式变压器的故障诊断过程中有一定的实际应用效果,但是该方法的精度完全取决于模型训练过程中的数据样本量,样本量越大预测精度越高,样本量越小预测精度越低,这对于少样本数据的分析过程是无法保障的,为此本文提出了适用于小样本数据的基于支持向量机集合的油浸式变压器故障诊断建模仿真研究方法。

2 支持向量机集合

2.1 支持向量机概念

支持向量机是机器学习方法在传统统计理论的应用成果,唯一的差别是传统统计理论遵循的是经验风险最小,而支持向量机遵循的是结构风险最小。该方法最大的特点就是适用于小样本数据,而神经网络适用于大样本数据,所以对于变压器故障诊断这种小样本数据,支持向量机的适用性更强^[3]。

支持向量机采用的是3层结构,中间隐层节可以根据不同的实际问题而做出变化,从而扩大了适用范围。通过支持向量机的核函数可以提高训练数

第一作者简介:肖海燕,女,工程师,本科,从事低压配电方向研究。
收稿日期:2020-05-18; 修回日期:2020-12-04

据的维数,将非线性问题的求解引入到高维空间,这样既可以保障不同实际问题的样本数据不受维数的限制,同时可以避免神经网络算法的维数短板。

2.2 支持向量机集合

支持向量机是针对二分类问题提出来的,但是变压器故障类型多种多样,不符合直接使用条件,为此选用一对一分类方法,核心思想是将多分类问题一一分解,成为多个二分类问题的组合,这样的计算过程简单明了,物理意义更加明确。变压器输入数据样本每组有 5 个参数,可表示为样本 (x, y, z, m, n) 。此类数据是线性不可分的,因此需要引入核函数,将二维空间内线性不可分的数据映射到一个高维特征空间内,巧妙地解决了数据在高维特征空间中的内积运算,从而很好地解决非线性分类问题。本文中选用的是运用最广泛的高斯径向基核函数。

支持向量机(SVM)集合的工作原理是:输入数据样本共有 n 个类别,每次选取其中的 2 个类别,对该 2 个类别的组合构造 SVM,总共需要构造 $n(n-1)/2$ 个 SVM,在本文中 $n=5$ 。在确定某个样本所属类别时,采用的是“投票法”,样本每经过一个二分类 SVM,都会输出 1 个可能的所属类别,即该类别获得 1 票。当样本经过所有二分类 SVM 后,对各个类别获得票数进行统计,得票数最高的类别即为该样本最可能属于的类别。当某几个类别获得相同票数时,再将样本输入与这几个类别相关的二分类 SVM,如此循环,直至最终出现票数胜出的类别^[4]。

3 基于支持向量机集合的故障诊断

3.1 输入输出变量确定

本次研究选取 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 和 C_2H_6 的体积分数作为输入变量,输出变量则选择变压器常见的 5 种运行状态。目前油浸式变压器常见的故障类型以电性和热性为主,电性故障可以进一步细分为低能放电故障和高能放电故障,热性故障可以进一步细分为 700 °C 以下的中低温过热故障和 700 °C 以上的高温过热故障。分别对状态进行编码,变压器运行状态对应编码和类别见表 1。

3.2 故障诊断流程

基于支持向量机集合的故障诊断流程见图 1。集合 C 为所有构造的支持向量机的集合。当从路径 A 进入二分类 SVM 选择器时,选择器选择集合 C 中所有的 10 个支持向量机,即样本经过 10 个支持向量机。当从路径 B 进入选择器时,若分类 1 和分

表 1 变压器运行状态对应编码和类别

变压器运行状态	编码	类别
正常	00001	1
低能放电	10000	2
高能放电	01000	3
中低温过热	00100	4
高温过热	00010	5

类 2 票数相同,则选择集合 C 中所有与分类 1 和分类 2 相关的 7 个支持向量机,即样本只经过这 7 个支持向量机。其余分类票数相同的情况与上述情况同理。

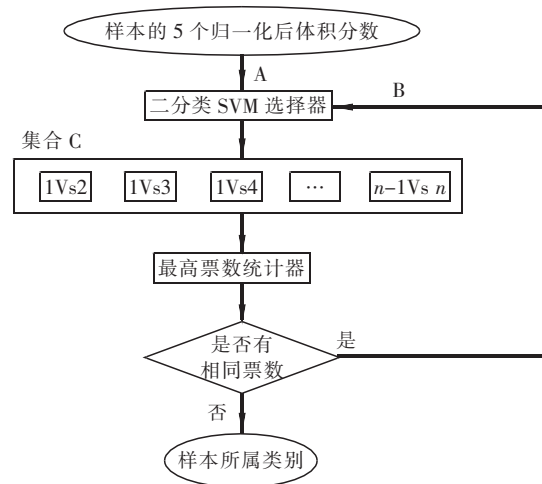


图 1 基于支持向量机集合的故障诊断流程

① 明确训练样本数据

诊断结果的有效性与样本数据的搜集质量密切相关,本次研究搜集了 240 组数据作为输入数据,按照 2:1 的比例划分为训练数据和测试数据,共有 160 组训练样本数据,80 组测试数据。每组数据为 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 和 C_2H_6 的体积分数。训练样本数据的样例见表 2。

表 2 训练样本数据样例

状态编码	H_2 体积分数 / 10^{-6}	CH_4 体积分数 / 10^{-6}	C_2H_6 体积分数 / 10^{-6}	C_2H_4 体积分数 / 10^{-6}	C_2H_2 体积分数 / 10^{-6}
00001	15	5	11	2	1
00001	24	11	6	9	0
10000	234	32	21	19	53
10000	121	35	10	75	15
01000	3 598	245	16	6	0
01000	4 023	289	12	4	0

续表2

状态编码	H ₂ 体积分数 /10 ⁻⁶	CH ₄ 体积分数 /10 ⁻⁶	C ₂ H ₆ 体积分数 /10 ⁻⁶	C ₂ H ₄ 体积分数 /10 ⁻⁶	C ₂ H ₂ 体积分数 /10 ⁻⁶
00100	45	87	47	123	0
00100	112	432	120	563	0
00010	134	167	32	168	13
00010	178	18	68	132	0

考虑到本次参与向量机运算的5种气体的体积分数差别较大,为了使各种气体体积分数在向量机运算过程中权值均衡,需对数据进行归一化预处理。论文采用的归一化方法是最大最小值归一化方法,见式(1),通过归一化预处理,使每一种气体体积分数的值都集中在0~1范围。以气体A为例,归一化方法如下。

$$\varphi_s = \frac{\varphi - \varphi_{\min}}{\varphi_{\max} - \varphi_{\min}} \quad (1)$$

式中 φ_s ——归一化后气体体积分数

φ ——样本气体A体积分数

φ_{\min} ——240组数据中气体A体积分数的最小值

φ_{\max} ——240组数据中气体A体积分数的最大值

② 选择核函数

核函数影响着数据从原始空间映射到高维空间的方式,所以核函数的选择是支持向量机建模过程中的重要环节。由于油浸式变压器的故障类别只有5种,所以本次研究选择了高斯径向基核函数的变体幂指数核函数^[5],其表达式为:

$$K(x, x_j) = \exp\left(-\frac{\|x - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

式中 K ——核函数

x ——输入特征变量

x_j ——第 j 个高斯径向基核函数的中心

σ ——高斯径向基核函数宽度

3.3 模型训练

首先利用训练样本数据对支持向量机模型进行训练,即建模过程。输入160组已知输出结果的样本数据,分类为1的样本数据只会进入所有与类别1相关的支持向量机,其他分类同理。通过程序对高斯径向基核函数宽度 σ 和扰动系数 A 进行全局搜索寻优,得到确定的 σ 和扰动系数 A 的取值范

围。 σ 反映了核函数的映射关系。 A 代表各种类型的误差允许范围,当训练样本噪声较大时,扰动系数 A 需要设置得较大,当训练样本噪声较小时,扰动系数 A 需要设置得较小。

3.4 模型测试

将80组测试数据输入模型,输出测试数据的所属类别。在80组样本数据的测试结果中,有69组预测结果正确,正确率达到了86.25%。经过分析,误差的原因主要有两个方面,一是支持向量机的核函数参数的选取没有达到最优,二是选取的数据样本之间并不是完全独立的,存在着相关性。

4 结论

油浸式变压器故障诊断是多分类问题。为了解决传统油浸式变压器故障诊断需要样本数据量大的问题,提出基于支持向量机集合的变压器故障诊断方法。将多分类问题分解为多个二分类问题的组合,变压器故障分为5个类别,构建10个支持向量机,每个支持向量机对应2个故障类别。样本每经过1个支持向量机都会获得1个可能的所属类别,当经过所有支持向量机后,通过投票法确定样本最终所属类别。通过模型训练,求得核函数的取值,建立模型。模型测试结果正确率达到86.25%,验证了方法的有效性,说明该方法适用于油浸式变压器故障诊断。

参考文献:

- [1] MIRANDA V, CASTRO A R G, LIMA S. Diagnosing faults in power transformers with auto-associative neural networks and mean shift[J]. IEEE Transactions on Power Delivery, 2012(3): 1350-1357.
- [2] 杨道武,李海如,杨世凤,等. 基于改进 PSO-BP 混合算法的电力变压器故障诊断[J]. 电力科学与技术学报, 2011(1): 99-103.
- [3] 郑含博,何乐,周浔,等. 基于多分类最小二乘支持向量机和改进粒子群优化算法的电力变压器故障诊断方法[J]. 高电压技术, 2014(9): 367-370.
- [4] 李赢,舒乃秋,古田均,等. 基于模糊聚类和完全二叉树支持向量机的变压器故障诊断[J]. 电工技术学报, 2016(9): 2104-2110.
- [5] 唐浩,屈梁生,付兴涛,等. 基于支持向量机的发动机故障诊断[J]. 西安交通大学学报, 2007(9): 91-96.

(本文责任编辑:李欣雨)