

基于 DGA 和 HMM 的小样本变压器故障诊断算法

刘敏^{1,2},王洪申¹,魏泰²

(1. 兰州理工大学 机电工程学院,甘肃 兰州 730050; 2. 甘肃省特种设备检验检测研究院,甘肃 兰州 730050)

摘要:针对基于油中溶解气体分析(DGA)的变压器故障诊断方法需要大量训练样本,而实际应用中缺乏样本的问题,提出在小样本情况下一种应用基于 DGA 和隐马尔科夫模型(HMM)的变压器故障诊断算法。该算法利用 DGA 提取变压器的故障特征,使用 HMM 作为变压器故障诊断分类器,并对 HMM 模型的训练算法进行改进,引入比例因子和多观测序列。试验结果表明:使用公开数据集 IEC TC 10 作为变压器故障数据集,将本算法与常用于小样本情况的 SVM、决策树算法相比较,本算法的变压器故障诊断正确率更高。

关键词:油中溶解气体分析;隐马尔科夫模型;B-W 算法;变压器;故障诊断

中图分类号:TM41 **文献标志码:**B **文章编号:**1671-5276(2023)04-0033-04

Small Sample Transformer Fault Diagnosis Algorithm Based on DGA and HMM

LIU Min^{1,2}, WANG Hongshen¹, WEI Tai²

(1. College of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;

2. Gansu Province Special Equipment Inspection and Testing Institute, Lanzhou 730050, China)

Abstract: To cope with the short supply of training samples needed in a great number for transformer fault diagnosis method based on dissolved gas analysis in oil (DGA), a transformer fault diagnosis algorithm based on DGA and hidden Markov model (HMM) is proposed. The algorithm uses DGA to extract transformer fault features, adopts HMM as transformer fault diagnosis classifier, improves the training algorithm of HMM model and introduces scale factor and multi observation sequence to improve the fault diagnosis accuracy. The experimental results show that with the application of the open data set IEC TC 10 as the transformer fault data set and in comparison with the SVM and decision tree algorithms commonly used in small samples, the accuracy of transformer fault diagnosis of the proposed algorithm is higher.

Keywords: dissolved gas analysis in oil; hidden Markov model; Baum-Welch algorithm; transformer; fault diagnosis

0 引言

电力变压器是电网输电环节中最关键的设备,及时发现变压器的潜伏性故障能有效保障电网的安全稳定运行,减少电力系统故障带来的经济损失。目前,国内外针对油浸式变压器故障诊断常用的检测方法是油中溶解气体分析 DGA(dissolved gas analysis)技术^[1-2],将变压器绝缘油中溶解的气体成分作为故障诊断的特征量,再结合 IEC 三比值法^[3]和罗杰斯比值法^[4]等传统方法来诊断故障。这种做法应用广泛,已积累了许多实践经验,我国现行的国家标准中使用的就是 IEC 三比值法^[5]。

传统方法虽然简单、快速,但仍存在诊断准确度不高和应用局限性等问题,以 IEC 三比值法为例,存在故障编码缺失的问题,导致实践中出现无法判断故障类型的情况^[6]。目前,已有新的智能诊断方法被引入变压器故障诊断研究领域,例如人工神经网络^[7]、模糊理论^[8]、支持向量机^[9]等。这些智能方法虽然故障诊断精度更高,但对训练样本数据集的完备性要求高^[10],训练算法需要大量带标签的样本,而实际应用中,难以满足此要求。

针对实际应用中缺乏数据样本的情况,本文引入无监督学习的隐马尔科夫模型(hidden markov model, HMM)^[11]来解决油浸式变压器故障诊断问题。HMM 是一种著名的有向图模型,已成功地用于语音识别^[12]、自然语言处理^[13]和故障诊断^[14]等领域。HMM 具有很强的学习能力和模式分类能力,能从训练数据中深入挖掘变压器故障的隐含信息,同时 HMM 的训练算法计算简单,收敛速度快,对样本数量的要求低,能应用在小样本的情况。因此本文将 HMM 作为分类器引入油浸式变压器故障诊断中,并对 HMM 的训练算法 B-W(baum-welch)^[15]进行改进,引入多观测序列和比例因子来提高算法的实用性。在小样本情况下,将本文算法与常用于小样本分类的 SVM 算法和其他算法相比较,试验证明本文算法具有更高的故障诊断精度。

1 HMM 算法概述

1.1 HMM 原理

HMM 描述了一个由隐状态组成的马尔科夫链,可以

由一个三元组 $(\pi, \mathbf{A}, \mathbf{B})$ 表示,其中 π 是初始状态概率向量, \mathbf{A} 是状态转移概率矩阵, \mathbf{B} 是观测概率矩阵。每个隐状态生成一个观测值,隐状态之间转移的概率构成一个状态转移概率矩阵 \mathbf{A} ,状态生成观测值的概率组成观测概率矩阵。状态的变换用一个状态序列描述,同理状态序列对应的观测值用一个观测序列表示,如图 1 所示。



图 1 HMM 原理

HMM 可以解决三类问题:概率问题、学习问题和预测问题^[15]。将 HMM 应用于变压器故障诊断问题时,需要将 HMM 作为分类器,这归属于学习问题。即将 DGA 数据作为观测序列,需要训练出得到最大观测概率的 HMM。即已知观测序列 O ,估计 HMM 的 $(\pi, \mathbf{A}, \mathbf{B})$ 参数。根据变压器故障数据的特点,本文拟使用离散的左右型 HMM。

1.2 HMM 的训练

B-W 算法是基于极大似然估计的无监督算法,用于训练 HMM,对各参数进行迭代计算。在实际应用时, B-W 算法存在两个问题:

观测序列问题:经典的 B-W 算法是单观测序列的,对于左右型的 HMM,不可能只用一个观测序列训练模型,为了有足够的对模型参数进行可靠估计,必须使用多个观察序列;

精度溢出问题:HMM 中参与计算的是小于 1 的各项概率值,计算前向概率和后向概率 $\alpha_t(i)$ 和 $\beta_t(i)$ 需要累乘各状态转移概率 a_{ij} 和观测概率 $b_j(i)$,实际计算时,概率值可能会趋于 0 而超过机器浮点数的精度范围,造成精度损失。在多观测序列和长观测序列情况下,问题更为明显。

为了解决上述问题,本文拟使用带比例因子的多观测序列的 B-W 算法训练 HMM。在计算过程中加入一个放大因子并在后续计算中消去,即可解决精度溢出问题。

B-W 算法计算公式如下:

$$\bar{a}_{ij} = \frac{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \hat{\alpha}_t^l(i) a_{ij} b_j(O_{t+1}^l) \hat{\beta}_{t+1}^l(j)}{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \hat{\alpha}_t^l(i) \hat{\beta}_t^l(i) / c_t^l} \quad (1)$$

$$\bar{b}_j(k) = \frac{\sum_{l=1}^L \sum_{t=1, O_t^l=v_k}^{T_l-1} \hat{\alpha}_t^l(i) \hat{\beta}_t^l(i) / c_t^l}{\sum_{l=1}^L \sum_{t=1}^{T_l-1} \hat{\alpha}_t^l(i) \hat{\beta}_t^l(i) / c_t^l} \quad (2)$$

式中: O^l 表示第 l 个观测序列; T_l 表示对应观测序列 O^l 的训练时序; N 是隐状态数目; $1 \leq i, j \leq N$; $\hat{\alpha}_t(i)$ 和 $\hat{\beta}_t(i)$ 是 t 时刻的前向概率和后向概率; c_t 是对应观测序列 l 的比例因子; a_{ij} 是状态 j 向状态 i 转移的概率; $b_j(O_t)$ 是观测序列 O 在 t 时刻对应观测值 j 的观测概率。前向概率 $\hat{\alpha}_t(i)$ 和后向概率 $\hat{\beta}_t(i)$ 计算公式如下:

$$\begin{cases} \hat{\alpha}_t(i) = \sum_{j=1}^N \hat{\alpha}_{t-1}(j) a_{ji} b_i(O_t) \\ c_t = 1 / \sum_{j=1}^N \hat{\alpha}_t(j) \\ \hat{\alpha}_t(i) = c_t \ddot{\alpha}_t(i) \\ \hat{\beta}_t(i) = \sum_{j=1}^N a_{ji} b_j(O_{t+1}) \hat{\beta}_{t+1}(j) \\ \hat{\beta}_t(i) = c_t \ddot{\beta}_t(i) \end{cases} \quad (3)$$

式中 $\ddot{\alpha}_t(i)$ 和 $\ddot{\beta}_t(i)$ 是中间变量。

利用上述公式可以对 HMM 模型进行训练。HMM 的训练流程如图 2 所示。

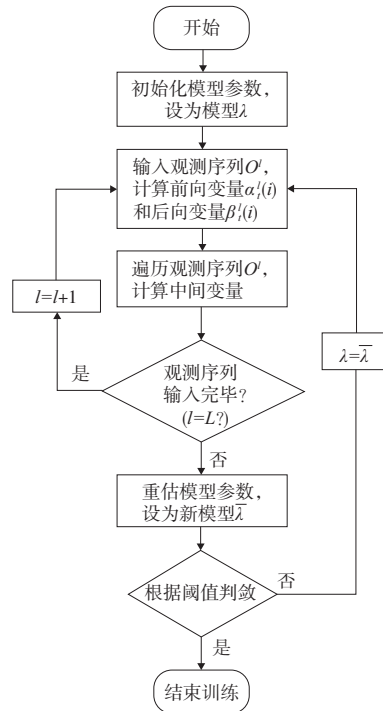


图 2 HMM 训练流程图

2 基于 HMM 的变压器故障诊断

2.1 变压器的故障分析

油浸式变压器的主要故障原因可分为电性和热性的,因为无运动部件,机械性的故障极少。发生故障时,变压器的绝缘材料会在电场或热场的影响下发生分解、老化等现象,产生氢气、二氧化碳、甲烷、乙烯、乙炔等气体并溶解在绝缘油中。故障类型不同,产生的气体种类和浓度都不同。通过 DGA 分析油浸式变压器的绝缘油就能获得变压器的故障信息。检修部门会按照规程对油浸式变压器进行定期的油样检测分析并将油样数据存储在数据库中,通过对记录的 DGA 数据进行分析,就能对油浸式变压器的故障进行诊断,及时发现潜伏性故障,为电力检修提供指导。参考 IEC60599 标准和国家标准 DL/T 7252—2014《变压器油中溶解气体分析和判断导则》^[5],本文将油浸式变压器故障类型分为中低温过热、高温过热、低能放电、

高能放电和局部放电 5 种。根据国家标准中的推荐做法, DGA 特征气体选取 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 、 C_2H_6 5 种。故障类型及特征气体如表 1 所示。

表 1 油浸式变压器故障类型及特征气体

样本编号	特征气体体积分数/(mm/L)					故障类型
	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6	
样本 1	3 420	7 870	33	6 990	1 500	中低温过热
样本 2	290	966	57	1 810	299	高温过热
样本 3	543	120	1 880	411	41	低能放电
样本 4	2 850	1 115	3 675	1 987	138	高能放电
样本 5	37 800	1 740	8	8	249	局部放电
样本 6	66	111	15	110	90	正常

2.2 数据预处理

油浸式变压器的 DGA 数据不能直接输入 HMM, 需要对其进行预处理。处理方法为先将数据进行归一化, 转换为 0~1 之间的向量形式, 再将其排序编码, 转换为离散 HMM 的输入形式。

归一化的公式为

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

排序编码的方法为: 根据故障种类确定划分区间的数目 N , 将归一化后特征向量的每一维按大小顺序排列并均等划分为 N 个区间, 在第 1 区间的全部编码为 0, 第 2 区间的全部编码为 1, ..., 在 N 区间的全部编码为 $N-1$ 。这样就将 DGA 数据转化为了 HMM 的输入向量形式, 以表 1 的数据为例, 处理结果如表 2 所示。

表 2 不同故障类型数据的归一化值和编码值

样本编号	特征气体的归一化值					编码值	故障类型
	H_2	CH_4	C_2H_2	C_2H_4	C_2H_6		
样本 1	0.43	1.00	0.00	0.89	0.19	(2,5,0,4,1)	中低温过热
样本 2	0.13	0.52	0.00	1.00	0.14	(1,3,0,5,1)	高温过热
样本 3	0.27	0.04	1.00	0.20	0.00	(2,0,5,2,0)	低能放电
样本 4	0.77	0.28	1.00	0.52	0.00	(4,2,5,3,0)	高能放电
样本 5	1.00	0.05	0.00	0.00	0.01	(5,1,0,0,0)	局部放电
样本 6	0.53	1.00	0.00	0.99	0.78	(3,5,0,5,4)	正常

2.3 HMM 分类器的使用

将 HMM 算法应用到变压器的故障诊断中可以分为预处理、模型训练和分类器的使用三步, 流程如图 3 所示, 具体描述如下。

预处理: 首先, 需要建立 HMM 模型, 对参数 (π, A, B) 初始化, 然后对变压器故障数据集进行预处理, 归一化和排序编码后转化为 HMM 模型的观测序列数据集, 再将数据划分为训练集和测试集。

模型训练: 使用带比例因子的多观测序列 B-W 算法

对 HMM 进行训练, 重估模型参数。使用特定类的样本训练 HMM, 可以得到对应的子分类器, 用 N 类故障样本训练 HMM 模型, 可以得到 N 个 HMM 子分类器。将这些子分类器组合在一起, 构成 HMM 分类器, 就可以对 N 类故障进行故障诊断。

分类器的使用: 使用 HMM 分类器时, 将样本输入训练好的 HMM 分类器中, 计算输入的观测序列 O 在每个子分类器中的概率值 $P(O|\lambda)$, 将样本归入取得概率最大的那个子分类器的故障类型中, 完成分类。

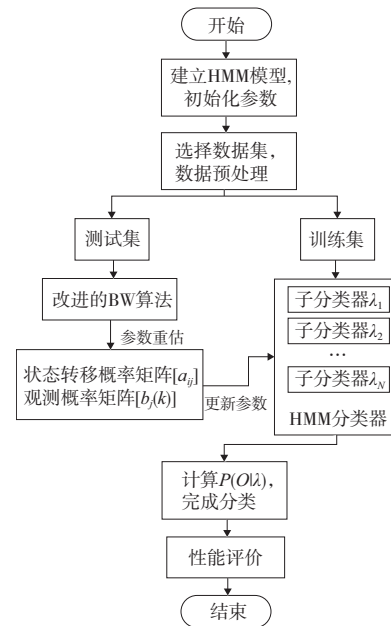


图 3 HMM 分类器的使用

3 试验及结果分析

本文从 IEC TC 10 数据库^[16]中选取油浸式变压器的 DGA 数据组成故障诊断数据集。故障类型选取中低温过热、高温过热、低能放电、高能放电和局部放电 5 种, 加上正常类型共 6 类, 特征气体选取 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 、 C_2H_6 。在统计学领域, 每一类样本的数据量小于 30 可称为小样本, 本文按照小样本的要求, 从每种类型的故障样本选取 30 组数据参与试验。试验模拟软件为 Matlab2020b, 硬件平台内存为 8 G, CPU 为 i5 4570。

为检验算法的故障诊断能力, 使用全部数据作为训练集, 将本文算法和常用于小样本的量机 SVM 相比较, 以决策树和朴素贝叶斯算法为对比补充。其中 HMM 模型的隐状态数目设为 6, 观测值数目设为 6, 初始状态矩阵、初始状态矩阵 A 和观测概率矩阵 B 均随机生成, 诊断结果如表 3 所示。图 4 是 3 类故障情况的 ROC 曲线 (receiver operating characteristic curve) 图 (因篇幅所限, 只列出 3 类情况)。从表 3 可以看出以 HMM 作为分类器, 不同故障的诊断正确率均高于 SVM 算法、决策树算法和朴素贝叶斯算法, 从图 4 的 ROC 曲线图可以看出, 3 种情况下的 HMM 的 AUC (area under curve) 值均为最大, 这说明 HMM 分类器的性能最优, 对故障的识别能力最强。综合

所述,本文算法在小样本的情况下诊断准确率和作为分类器的性能都常用于小样本的 SVM 等算法高,具有一定的优势。

表 3 3 类算法的诊断正确率比较 单位:%

项目	不同算法的诊断正确率			
	HMM	SVM	决策树	朴素贝叶斯
中低温过热	80.0	53.3	60.0	40.0
高温过热	90.0	86.7	83.3	25.0
低能放电	90.0	70.0	43.3	16.7
高能放电	86.7	73.3	66.7	36.7
局部放电	90.0	86.7	83.3	25.0
正常	93.3	93.3	83.3	90.0
总体正确率	87.2	75.6	69.4	46.4

为了进一步检验算法泛化性能,并考虑到小样本情况下训练数据的划分随机性问题,使用留出法进行 10 次重复试验,试验结果取 10 次的平均值以减小随机性的影响。试验结果如表 4 所示。可以看出,随着训练集占数据集比例的增大,4 类算法的故障诊断正确率均有提高,其中 HMM 的正确率均为最高,说明了 HMM 的泛化能力比 SVM 算法、决策树算法和朴素贝叶斯算法更强。

表 4 3 类算法在不同训练集下的诊断正确率 单位:%

训练集划分比例	不同算法的诊断正确率			
	HMM	SVM	决策树	朴素贝叶斯
50	58.7	56.0	52.4	44.8
60	63.5	59.6	56.9	45.0
70	68.6	67.8	57.6	45.1
80	70.8	69.6	62.0	45.8

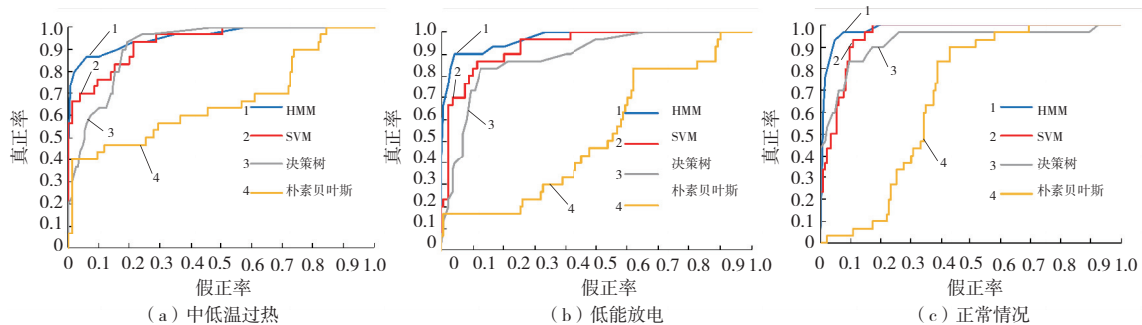


图 4 3 类故障情况下的 ROC 曲线图

4 结语

本文提出一种将 DGA 和 HMM 相结合、在小样本情况下使用变压器故障诊断算法。该算法对无监督的经典 B-W 算法进行改进,加入比例因子和多观测序列,并将其应用在 HMM 的训练中。将本文算法与常用于小样本情况下的 SVM 算法、决策树算法和朴素贝叶斯算法相比较,对公开数据集 IEC TC 10 数据库中的故障样本进行诊断,试验结果表明,本文算法具有更高的识别精度和更强的泛化能力。

参考文献:

- [1] 操敦奎. 变压器油色谱分析与故障诊断[M]. 北京:中国电力出版社,2010.
- [2] WANI S A, RANA A S, SOHAIL S, et al. Advances in DGA based condition monitoring of transformers: a review [J]. Renewable and Sustainable Energy Reviews, 2021, 149: 111347.
- [3] 高文胜, 严璋, 谈克雄. 基于油中溶解气体分析的电力变压器绝缘故障诊断方法[J]. 电工电能新技术, 2000, 19(1): 22-26.
- [4] ROGERS R R. IEEE and IEC codes to interpret incipient faults in transformers, using gas in oil analysis[J]. IEEE Transactions on Electrical Insulation, 1978, EI-13(5): 349-354.
- [5] DL/T 722—2014 变压器油中溶解气体分析和判断导则[S].
- [6] 宋斌, 刘志雄, 李恩文, 等. DGA 中三比值缺码研究初探[J]. 电力自动化设备, 2015, 35(12): 60-65.
- [7] YANG X H, CHEN W K, LI A Y, et al. BA-PNN-based methods for power transformer fault diagnosis[J]. Advanced Engineering Informatics, 2019, 39: 178-185.

- [8] 薛盛炜, 李川, 李英娜. 改进模糊聚类与主成分分析下的变压器故障识别[J]. 河南科技大学学报(自然科学版), 2020, 41(6): 39-44, 50, 5.
- [9] ZHANG Y Y, WEI H, YANG Y D, et al. Forecasting of dissolved gases in oil-immersed transformers based upon wavelet LS-SVM regression and PSO with mutation[J]. Energy Procedia, 2016, 104: 38-43.
- [10] 王国平, 余涛, 傅森木, 等. 基于 DGA 的变压器故障诊断智能方法分析[J]. 电力建设, 2015, 36(6): 34-39.
- [11] 李航. 统计学习方法[M]. 2 版. 北京:清华大学出版社, 2019: 174-184.
- [12] 周楠, 艾剑良. 基于 HMM 和 RNN 的无人机语音控制方案与仿真研究[J]. 系统仿真学报, 2020, 32(3): 464-471.
- [13] SMIT P, VIRPIOJA S, KURIMO M. Advances in subword-based HMM-DNN speech recognition across languages[J]. Computer Speech & Language, 2021, 66: 101158.
- [14] NTALAMPIRAS S, SOUPIONIS Y, GIANNPOULOS G. A fault diagnosis system for interdependent critical infrastructures based on HMMs[J]. Reliability Engineering & System Safety, 2015, 138: 73-81.
- [15] RABINER L R. A tutorial on hidden Markov models and selected applications in speech recognition[J]. Proceedings of the IEEE, 1989, 77(2): 257-286.
- [16] DUVAL M, DEPABLA A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases[J]. IEEE Electrical Insulation Magazine, 2001, 17(2): 31-41.

收稿日期: 2021-12-13