

基于朴素贝叶斯算法的电力变压器故障诊断方法研究

改良三比值法是我国推荐的变压器故障诊断方法，该方法在预试试验和实验室分析中具有很高的正判率。然而由于DGA在线监测数据的准确性不高造成改良三比值法正判率下降，诊断结果难以支撑变压器智能化的发展要求。因此，提出了一种基于朴素贝叶斯的DGA故障诊断算法。该算法对数据进行预处理后，采用由三个条件属性构成的三元组来描述DGA数据特征，并通过朴素贝叶斯算法计算训练样本的先验概率和测试样本的后验概率，从而进行故障诊断。通过实测数据集上的验证表明，该算法识别正判率优于改良三比值法，且具有可行性和有效性。

雍明超 吕 侠 周 钟 牧继清 毛丽娜/许继集团有限公司

0 引言

电力变压器是电力系统运行的主要组成部分，其安全稳定运行是保证电力系统稳定的重要条件。在热、电等的作用下，变压器运行中的油纸绝缘材料会发生老化裂解，产生低分子烃类和 CO_2 、 CO 等气体溶于油中。因此对油中溶解气体的监测与分析就可以间接反映变压器的绝缘故障。我国自开展变压器在线监测建设以来，油中溶解气体分析(Dissolved Gas Analysis, DGA)就作为变压器在线监测的主要内容，积极开展基于故障气体的研究工作^[1-3]。据统计，在出现故障的变压器中，过热性故障为63%，高能放电故障占18.1%，过热兼高能放电占10%，火花放电占7%，受潮故障占1.9%。而基于DGA数据的故障分析能够相对准确地识别上述故障，基于此国家电网发布的一系列导则规范确定DGA作为新建110 kV变电站的必要配置。DGA的监测数据和就地诊断结果是

检修的重要参考意见，这对基于DGA的变压器故障诊断算法的准确性提出了更高的要求。

基于DGA的变压器故障诊断算法一般采用基于规则的方法，国内外经常应用的有立方图法、电协研法、四比值法及大卫三角形法等^[4-5]。特别是操敦奎等学者在大量的试验统计之后提出的改良三比值法，成为我国基于DGA数据的变压器故障诊断的行业标准。但是，基于规则的诊断方法对监测数据的完整性及准确性有较高要求，边界数据处理相对困难，并且不允许监测数据缺失。

近年来，一些专家基于DGA数据利用智能算法进行变压器故障诊断的研究。文献[6-7]运用模糊理论和粗糙集进行变压器故障诊断，该方法具有结构简单和诊断速度快的优点，但学习能力较差。文献[8-10]描述了神经网络算法在变压器故障诊断的应用，该算法的优点是自学能力强且有并行处理能力，但收敛速度较慢。贝叶斯算法在分布式电网故障诊断、电能质量监测、配电网可靠性诊断、变压器可靠性分析、变压器综合诊断以及机械系统可靠性分析中都有良好的应用^[11]。贝叶斯算法是根据事件的后验概率求取事件的后验概率，然后根据后验概率大小对事件进行分类。它具有结构简单、自学习及运算速度快的特点。因此，本文提出了一种基于DGA在线监测数据的朴素贝叶斯的变压器故障诊断算法。该算法基于DGA的监测数据对数据集进行预处理，采用由三个条件



雍明超/工程师

关键词/Keywords

油中溶解气体分析·
改良三比值·
朴素贝叶斯·
故障诊断·

属性构成的三元组来描述DGA数据特征,并形成训练样本和测试样本。通过朴素贝叶斯算法计算训练样本的先验概率和测试样本的后验概率,从而进行故障诊断。验证表明,该算法优于改良三比值法,且具有可行性和有效性。

1 基于DGA的变压器诊断算法

1.1 变压器内部故障与油中气体含量关系

目前,DGA在线监测通常采用气相色谱法原理,通过进油、脱气将气体吹扫到色谱柱中进行分离,最终分析出 H_2 、 CH_4 、 C_2H_2 、 C_2H_4 、 C_2H_6 、 CO 和 CO_2 等故障气体。气体的种类及其含量间的关系与变压器的故障类型有着密切的联系,根据故障气体可以间接反映故障类型。

目前改良三比值法是比值法中应用最广的算法。据文献[1]对671台变压器的统计分析表明,在实验室条件下正判率可以达到97.1%。国内行业标准和国标也将改良三比值法作为变压器内部故障诊断的主要方法。但是在线监测数据的准确性与实验室数据相比有很大的差异,监测的准确度误差在30%即满足准确度要求,造成诊断的正判率下降,诊断结果难以支撑变压器智能化的发展要求。因此,研究更有效的算法符合变压器智能化的发展方向,具有重要的意义。

1.2 朴素贝叶斯分类器

贝叶斯分类方法建立在数学中的统计学基础上,根据已有的样本数据实例,运用贝叶斯公式来判定待测的数据归类问题。贝叶斯方法的主要思想是将事件的先验概率和后验概率相联系,并利用先验信息预测事件的后验概率。贝叶斯分类器具有自学习的能力,可以将新数据记录加入到已知的先验信息中,从而进一步影响事件的后验概率,提高事件预测的准确性。

贝叶斯公式如下

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (1)$$

式中, $P(C|X)$ 为条件 X 下 C 的后验概率; $P(C)$ 为 C 的先验概率; $P(X|C)$ 为条件 C 下的后验概率; $P(X)$ 为

X 的先验概率。

令 A 代表条件属性变量集, C 代表故障类型变量。假设有 n 个条件属性变量, $Val(A_i)$, $1 < i < m$ 代表条件属性变量的取值范围, $Val(C)$ 代表故障类型变量的取值范围,则 $A = \{A_1, A_2, \dots, A_n\}$, A_i 的取值 $a_i \in Val(A_i)$, C 的取值 $c \in Val(C)$ 。 t 代表训练样本, $t = \{a_1, a_2, \dots, a_n, c_l\}$, $1 < l < m$ 表示训练样本实例, x 表示测试样本, $x = \{a_1, a_2, \dots, a_j, \dots, a_n\}$ 表示测试样本实例。

样本 x 归属于类 c_j 的后验概率

$$\begin{aligned} P(c_j | a_1, a_2, \dots, a_n) &= \frac{P(a_1, a_2, \dots, a_n | c_j)P(c_j)}{P(a_1, a_2, \dots, a_n)} \quad (2) \\ &= \alpha P(c_j)P(a_1, a_2, \dots, a_n | c_j) \end{aligned}$$

式中, $P(c_j | a_1, a_2, \dots, a_n)$ 为测试样本归入故障类型 c_j 的后验概率; $P(a_1, a_2, \dots, a_n | c_j)$ 为故障类型 c_j 的条件概率; $P(c_j)$ 为故障类型 c_j 的先验概率; $\alpha = 1 / P(a_1, a_2, \dots, a_n)$, α 为常量。

贝叶斯算法根据式(2)计算出全部类 c_j , $j = 1, 2, \dots, m$ 的后验概率,然后将该组测试样本实例归入后验概率最大的类中。贝叶斯分类器将因果关系和概率知识结合在一起,符合人们的常性思维,效果良好,在实际中有广泛的应用。朴素贝叶斯分类器是最早的贝叶斯分类器,以各属性变量间相互条件独立为基础,即假设某一属性取值的概率完全不受其他属性的影响。由于属性间的相互独立,则式(1)中的后验概率的计算可变换为

$$P(c_j | a_1, a_2, \dots, a_n) = \alpha P(c_j) \prod_{i=1}^n P(a_i | c_j) \quad (3)$$

贝叶斯分类器具有良好的描述系统事件多态性和各事件之间逻辑关系非确定性的性能,并能进行正反向概率推理,同时贝叶斯分类器可直接求得系统的故障率、根节点的后验概率等,其优越的状态描述和推理机制使贝叶斯分类器在复杂系统的可靠性分析领域得到了广泛应用^[6-9]。

2 基于朴素贝叶斯的DGA诊断算法

2.1 特征集提取

朴素贝叶斯的运算需要构建特征数据集。改

良三比值法的三个比值规则在比值法中具有代表性,朴素贝叶斯特征集的每个属性变量的具体取值也取自这三个比值。

算法特征集用一个三元组 V 表示,即

$$V = \{C_2H_2 / C_2H_4, CH_4 / H_2, C_2H_4 / C_2H_6\} \quad (4)$$

2.2 故障诊断算法原理

算法首先搜集溶解气体的相关信息,并进行数据处理,得到该算法所需的特征集,同时形成训练样本集和测试样本集。训练样本集 $T = \{A, C\}$ 。其中 A 表示条件属性集,即 $A = \{A_1, A_2, A_3\}$, $Val(A_1) = Val(A_2) = Val(A_3) = \{0, 1, 2\}$; C 表示类型属性, $Val(C) = \{\text{低能放电}, \text{高能放电}, \text{低能放电兼过热}, \text{低温过热}, \text{中温过热}, \text{高温过热}, \text{高能放电兼高温过热}\}$ 。测试样本集中类型属性格式与训练样本集一致,但测试样本集中的故障类型 C 为空。算法流程如图1所示。

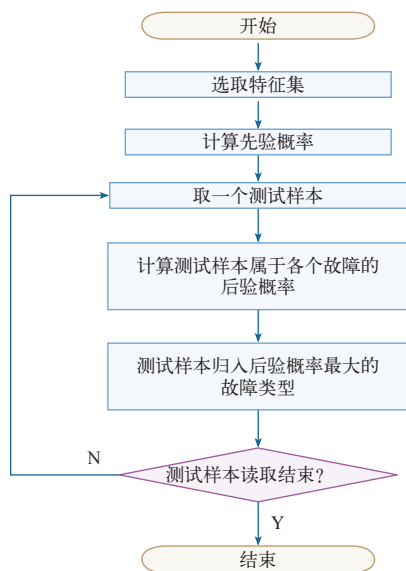


图1 算法流程图

3 算例分析

3.1 算例数据集

算例数据收集自实际运行的电力变压器DGA在线监测数据,变压器涵盖110~500 kV电压等级。样本数据集中包含525个完整样本。数据样本的故障类型分布如下表所示。

表 样本的故障类型分布

故障类型	数据集样本数量
低能放电	137
低能放电兼过热	29
高能放电	91
低温过热	4
中温过热	33
高温过热	192
高能放点兼高温过热	39

3.2 算例结果及分析

推理该分类器范例检索过程中的特性指标,选取 H_2 、 CH_4 、 C_2H_6 、 C_2H_4 、 C_2H_2 、 CO_2 和 CO 七种特征气体,搜索算法采用基于Pearson积矩相关系数的范例检索算法。

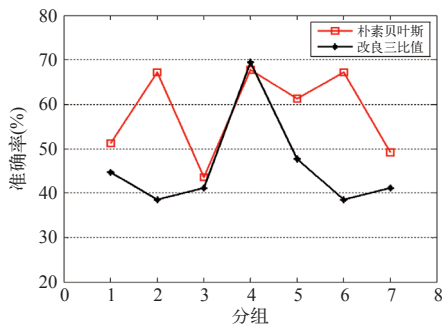
源范例 x 和目标范例 y 间的Pearson积矩相关系数可表示为

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5)$$

式中, X_i 、 Y_i 分别为范例 x 和 y 的第 i 个特性指标; $\bar{X}$ 、 $\bar{Y}$ 分别为范例 x 和 y 的均值。

相关系数 r 用来描述范例间的相关程度,为 $-1 \sim 1$ 之间的数, r 值越大,目标范例和源范例越接近。取1时,表示变量 X 和 Y 之间具有线性变化的关系,即 Y 随着 X 的增加而增加,而且所有的点都落在一条直线上;取 -1 时,则所有点落在一条直线上,但是变量 Y 随着 X 的增加而减小。相关系数值为0时,表示变量之间没有线性相关关系。

图2给出了本文算法和三比值法在数据集上的测试结果。



(a) 第一次十折交叉验证结果

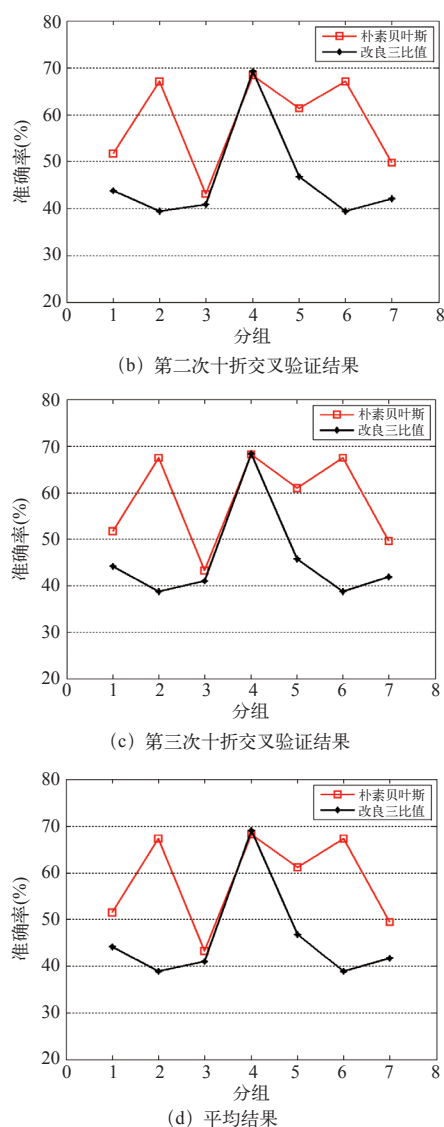


图2 朴素贝叶斯算法和三比值法的对比结果

上述验证结果表明：

1) 朴素贝叶斯在6个分组上正判率明显优于改良三比值法，只在低温过热故障类型中二者正判率相当。

2) 基于朴素贝叶斯的DGA诊断算法是可行的、有效的。

4 结束语

针对目前DGA在线监测存在较高误差造成普遍采用的改良三比值法正判率低下的情况，本文基于统计的方法提出了一种基于朴素贝叶斯的DGA诊断算法。该算法首先采用三元组表示数据特征，然后运用贝叶斯公式对待测样本进行分

类。算例试验表明该算法整体正判率优于改良三比值法，可以适用于DGA在线监测就地预警，满足分级诊断的要求，具有较大的工程实用价值。

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，2010：83-88.
- [2] 操敦奎，许维宗，阮国方. 变压器运行维护与故障分析处理 [M]. 北京：中国电力出版社，2008：156-170.
- [3] 赵峰，李硕，等. 基于DGA和改进型灰关联度模型的牵引变压器故障诊断 [J]. 高压电器，2015，51(1)：41-45.
- [4] 阎春雨. 采用油中溶解气体分析法判断变压器故障应注意的事项 [J]. 变压器，2006，43(9)：38-41.
- [5] 杨廷方，李景禄，曾祥君，等. 基于多方法组合诊断模型的大型变压器故障诊断 [J]. 电力系统自动化，2009，33(20)：92-95.
- [6] 武中利，杨建，朱永利，等. 基于粗糙集理论和支持向量机的变压器故障诊断 [J]. 电力系统保护与控制，2010，38(18)：80-83.
- [7] 公茂法，张言攀，柳岩妮，等. 基于BP网络算法优化模糊Petri网的电力变压器故障诊断 [J]. 电力系统保护与控制，2015，43(3)：113-117.
- [8] 程声烽，程小华，杨露. 基于改进粒子群算法的小波神经网络在变压器故障诊断中的应用 [J]. 电力系统保护与控制，2014，42(19)：37-42.
- [9] 张蕊，韩超，李晓娜. 基于IPSO-BP神经网络的变压器故障诊断方法 [J]. 计算机系统应用，2013，22(4)：125-126.
- [10] 程加堂，熊伟，段志梅，等. 灰色神经网络在变压器故障诊断中的应用 [J]. 高压电器，2010，46(8)：50-56.
- [11] 霍利民，朱永利，张在玲，等. 贝叶斯网络在配电系统可靠性评估中的应用 [J]. 电工技术学报，2004，19(8)：113-118.

(收稿日期：2016-01-04) EA