

文章编号:1001-1609(2006)01-0011-03

基于改进模糊 ISODATA 算法的变压器故障诊断

王子建, 何俊佳, 尹小根

(华中科技大学电气与电子工程学院, 湖北 武汉 430074)

Fault Diagnosis of Transformer Based on Revised Fuzzy ISODATA Algorithm

WANG Zi-jian, HE Jun-jia, YIN Xiao-gen

(College of Electrical & Electronic Engineering, Huazhong University of Science & Technology, Wuhan 430074, China)

摘要: 模糊 ISODATA 算法在基于变压器 DGA 的故障诊断中存在一些问题,如:模式空间的划分缺乏依据,聚类分析时没有考虑各种气体成分对故障反映的灵敏度等。笔者对此进行了改进,引入了一个描述不同气体成分对故障反映灵敏度的指标权向量,并在每次迭代运算之后对聚类中心进行分解和合并处理。利用改进的 ISODATA 算法对 3 起变压器故障进行了分析,得到了比较高的判断准确度。

关键词: 变压器; 故障诊断; 溶解气体分析; 聚类分析

中图分类号: TM855

文献标识码: A

Abstract: The fuzzy ISODATA algorithm has some disadvantages in the transformer fault diagnosis based on DGA, such as, lack basis of dividing the mode space, and not considering the reflection to the sensitivities of various gases to fault and so on. Some aspects of revision are proposed, i.e., a weight vector of index is introduced to describe the reflection to the sensitivities of different gases to fault; the cluster centers are split and merged after each iterative calculation. Three groups of transformer fault data are analyzed by using the revised algorithm, and the result shows that it has higher accuracy.

Key words: transformer; fault diagnosis; dissolved gas analysis (DGA); cluster analysis

1 引言

聚类分析的基本原理是在没有先验知识的情况下,基于“物以类聚”的观点,用数学工具分析各样本向量之间的距离及分散情况,按照样本的距离远近划分类别^[1]。

目前常用的一些聚类算法主要有:C-均值算法、迭代自组织数据分析算法(ISODATA)和基于样本和核相似性度量的动态聚类算法等。文[2]将模糊 ISODATA 算法应用到变压器油中溶解气体分析中来,并取得了较好的效果,但该算法在实际应用中还

存在着以下不足。

首先,该算法在聚类分析之前,需要先根据经验或借鉴其他故障诊断方法的分类来确定故障的类型,然后再将故障样本按照一定的规则分配到各个故障类型中去,最后再用聚类分析的方法对这种人为构造的模式空间与类型空间的映射关系进行分析。这种强加在故障样本与故障类型之间的对应关系缺乏对样本数据的科学分析,还不足以体现模式空间的真正结构。因此,在确定故障类型之前,有必要对故障样本数据进行科学的分析。

其次,ISODATA 算法为了推导迭代公式,引入了指数 β 。指数 β 的引入,在 β 值的定量上增加了随意性, β 值太小会引起计算溢出, β 值太大则会导致失真,且其数学物理概念不清晰,应用时可导致聚类结果不准确,甚至不可信^[3]。

另外,该算法没有考虑不同气体成分对于分类的不同影响,而实际故障诊断中,不同气体对于判别故障类型的影响是不同的,也就是说,不同气体成分对故障反映的灵敏度不同,如,乙炔就是一种对故障反映比较敏感的气体,因此有必要将各种气体成分的不同权重引入到聚类准则函数中来。

针对以上问题以及变压器油中溶解气体数据本身的一些特点,笔者对传统的模糊 ISODATA 算法提出以下几点改进:①在每次迭代运算之后,对聚类中心进行分解和合并处理;②将指标权向量 W 引入到聚类准则函数中来。利用该改进的 ISODATA 算法对变压器故障样本进行了聚类分析,得到了较好的效果。

2 故障样本数据的模糊化预处理

由于故障气体中各种气体的含量悬殊很大,另外,由于油中溶解气体的含量跟故障持续的时间、故障部位的面积、采样时变压器油的温度以及采样点

收稿日期:2005-08-04

作者简介:王子建(1981-)男,硕士研究生,主要从事电气设备在线监测与故障诊断的研究。

位置的选取等诸多因素有关，所以即使对于同一类故障，有时候所对应的故障气体的含量在数量级上也会有很大的差别。因此在聚类分析之前需要对故障样本数据进行一定的预处理。

目前，常用的预处理方法主要有：均值方差化、正规化变换、范数标准化、极大值规格化和比例规格化等，文[4]中还提到了先进行比例规格化，再用反正切函数作为隶属函数的处理方法。笔者通过比较分析，决定采用先比例规格化再正规化变换的预处理方法。

设待聚类的 n 个故障样本组成样本集合 $X=(x_1, x_2, \dots, x_n)$ ，每个样本用 m 个指标特征值向量表示： $x_i=(x_{i1}, x_{i2}, \dots, x_{im})$ ，求取隶属度的预处理算法如下：

$$x_{ij}' = x_{ij} / \sum_{j=1}^m x_{ij}$$
$$x_{ij}'' = \frac{x_{ij}' - \min_{1 \leq i \leq n} \{x_{ij}'\}}{\max_{1 \leq i \leq n} \{x_{ij}'\} - \min_{1 \leq i \leq n} \{x_{ij}'\}}$$

式中： $i=1, 2, \dots, n; j=1, 2, \dots, m$ 。则可用 $n \times m$ 阶指标特征值规格化矩阵：

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nm} \end{bmatrix} = (r_{ij})$$

对故障样本进行聚类分析。

3 改进的模糊 ISODATA 算法

该算法主要分为以下几个步骤：

(1) 选取起始聚类中心及控制参数。关于起始聚类中心的选取，文[5]中给出了 6 种方案，笔者采用以下方案： $\nu_1=(\bar{r}_1, \bar{r}_2, \dots, \bar{r}_m)$ ； $\nu_{2,3}=(\bar{r}_1 \pm s_1, \bar{r}_2, \dots, \bar{r}_m, \dots)$ ， $\nu_{c-1,c}=(\bar{r}_1, \bar{r}_2, \dots, \bar{r}_m \pm s_m)$ ，聚类中心的个数 $c=2m+1$ 。式

$$\text{中：}\bar{r}_j = \frac{1}{n} \sum_{i=1}^n r_{ij}; s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{ij} - \bar{r}_j)^2} \quad (j=1, 2, \dots, m)。$$

另外还有一些控制参数：分解参数 σ_0 ；合并参数 D_0 ；每一类中应该包含的最少样本个数 n_0 ，少于此数不能作为一个独立的聚类；最大运算次数 I_0 和迭代误差限 ε 。

(2) 运用迭代公式进行聚类分析。ISODATA 聚类迭代公式最早由贝狄克(Bezdek)提出，该算法聚类准则函数为：

$$\min\{F = \sum_{i=1}^n \sum_{k=1}^c [a_{ik}^\beta \sum_{j=1}^m (r_{ij} - \nu_{kj})^2]\}$$

式中： a_{ik} 为样本 i 归属于类别 k 的相对隶属度； ν_{kj} 为第 k 个聚类中心的第 j 个指标分量。

针对前面提到的该算法存在的不足，笔者采用文[3]的思想，引入指标权向量：

$$W = (w_1, w_1, \dots, w_m) = \left(\frac{s_1}{\sum_{j=1}^m s_j}, \frac{s_2}{\sum_{j=1}^m s_j}, \dots, \frac{s_m}{\sum_{j=1}^m s_j} \right) \quad (1)$$

其中： $s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (r_{ij} - \bar{r}_j)^2}$ ， $j=1, 2, \dots, m$ 。由该指标权向量建立聚类准则函数：

$$\min\{F = \sum_{i=1}^n \sum_{k=1}^c [a_{ik}^2 \sum_{j=1}^m (\omega_j (r_{ij} - \nu_{kj}))^2]\}$$

其数学物理意义为样本集对所有聚类中心的广义欧氏权距离平方和最小。利用拉格朗日乘数法得到迭代公式：

$$a_{ik}^{(l+1)} = 1 / \sum_{h=1}^c \frac{\sum_{j=1}^m [\omega_j (r_{ij} - \nu_{kj}^{(l)})]^2}{\sum_{j=1}^m [\omega_j (r_{ij} - \nu_{hj}^{(l)})]^2} \quad (2)$$

$$\nu_{kj}^{(l+1)} = \frac{\sum_{i=1}^n (a_{ik}^{(l)})^2 r_{ij}}{\sum_{i=1}^n (a_{ik}^{(l)})^2} \quad (3)$$

式中： $i=1, 2, \dots, n; j=1, 2, \dots, m; k=1, 2, \dots, c$ 。利用式(1)~式(3)进行迭代运算，直至满足条件： $\max\{|F = a_{ik}^{(l+1)} - a_{ik}^{(l)}|, 1 \leq i \leq n, 1 \leq k \leq c\} < \varepsilon$ 。得到聚类中心矩阵 V 和分类矩阵 A 。

(3) 聚类中心的分解。首先对聚类中心进行整理，如果某一类中的样本个数小于 n_0 ，则取消该聚类中心，且 $c-1$ ，然后，计算每类样本的标准差向量：

$$\sigma_k = (\sigma_{k1}, \sigma_{k2}, \dots, \sigma_{km}); \sigma_{kj} = \sqrt{\frac{1}{n_k} \sum_{i=1}^{n_k} (r_{ij} - \bar{r}_{kj})^2}$$

式中： n_k 为第 k 类中样本的个数； \bar{r}_{kj} 为第 k 类中第 j 个指标的平均值。

求每一个聚合类中各种指标标准差的最大值，以 $\{\sigma_{k\max}, k=1, 2, \dots, c\}$ 表示。如有 $\sigma_{k\max} > \sigma_0 [I]$ (给定)， I 为运算次数，则将聚类中心 ν_k 分解成两个新的聚类中心 ν_k^+ 和 ν_k^- ，且 c 加上 2。其中 ν_k^+ 相当于在 ν_k 的 $\nu_{k\max}$ 分量上加上 $k\sigma_{k\max}$ (k 给定)； ν_k^- 相当于在 ν_k 的 $\nu_{k\max}$ 分量上减去 $k\sigma_{k\max}$ 。

(4) 聚类中心的合并。计算全部聚类中心之间的距离

$$D_{ij} = \sqrt{\sum_{k=1}^m (\nu_{ik} - \nu_{jk})^2}, i=1, 2, \dots, c-1, j=i+1, \dots, c$$

将满足 $D_{ij} < D_0 [I]$ (给定) 的值按从小到大的顺序排列，然后按顺序对这些聚类中心进行合并，且 c 减去 1，得到新的聚类中心如下：

$$V_i^* = \frac{1}{n_i + n_j} [n_i \nu_i + n_j \nu_j]$$

(5) 如果是最后一次运算 (即第 I_0 次)，算法结束，否则 GOTO 第(2)步。

4 故障诊断实例分析

作者将收集到的 464 组变压器故障记录随机分为两份,各含 344,120 组记录,一份用于故障建模,一份用于准确率测试。笔者在 344 组故障记录的基础上,运用上述算法进行了聚类分析,最终形成聚类中心 32 个。在以后的故障诊断中,为了节省程序运行时间,可以直接应用这 32 个聚类中心作为起始聚类中心,运用迭代公式进行聚类分析。另外,在同一聚合类中为了找到与待诊断样本最相近的故障样本,采用了模糊理论中求取模糊贴近度的方法。设第 h 个聚合类中有 n_h 个样本,则样本 i 与样本 j 之间模糊贴近度的计算公式如下:

$$a_{ij} = 1 - f \sqrt{\sum_{k=1}^m \omega_k (r_{ik} - r_{jk})^2}; \quad \omega_k = \frac{\sqrt{\sum_{i=1}^{n_h} (r_{ik} - \bar{r}_k)^2}}{\sum_{k=1}^m \sqrt{\sum_{i=1}^{n_h} (r_{ik} - \bar{r}_k)^2}}$$

式中: \bar{r}_k 为聚合类中样本第 k 个指标的平均值, f 为给定参数。以下是运用改进的模糊 ISODATA 算法进行故障诊断的 3 个实例。

例 1: 四川电力局大水井站 1 号主变发生故障, DGA 数据如下: H_2 为 188, CH_4 为 236, C_2H_6 为 18.1, C_2H_4 为 237, C_2H_2 为 31.8 (单位: $\mu L/L$)。三比值编码: 122, 判断: 无对应编码。吊芯结果: C 相线圈分接头抽头引线对压铁马蹄口放电, 高压侧 A, B 相线圈对压环放电, 低压侧 B 相尾部烧焦。

该方法诊断结果见表 1。根据类中样本的故障类型, 将该例故障判断为高能放电兼过热, 与吊芯结果基本吻合。

表 1 聚类分析结果

类中样本	模糊贴近度	样本实际故障原因
160	0.924 501	高温过热
169	0.930 882	高于 700 °C 高温过热
183	0.925 893	高温过热
200	0.962 808	高温过热
261	0.962 433	高能放电
262	0.951 453	高能放电
308	0.976 003	电弧放电兼过热
310	0.962 560	高能放电兼过热
311	0.962 894	C 相高压侧引线周围纸板上有放电击穿和烧焦痕迹
312	0.965 478	绕组上部压紧螺钉的环氧压碗被压裂, 螺钉下端对铁压环放电并烧毛, 3 层绝缘板烧穿

例 2: 湖北电力局黄石电厂 2 号主变 (SFSL-60000/110) 发生故障, 各气体含量如下: H_2 为 93, CH_4 为 58, C_2H_6 为 43, C_2H_4 为 37, C_2H_2 为 0 (单位: $\mu L/L$)。三比值编码: 000, 判断: 无故障。吊芯结果: 上轭铁两点接地, 中温过热, 温度 510 °C。该方法诊断结果见表 2。

表 2 聚类分析结果

类中样本	模糊贴近度	样本实际故障原因
6	0.917 052	低温过热
60	0.999 996	中低温过热
201	0.857 684	高温过热
222	0.928 742	局部放电
232	0.965 798	低能量放电
240	0.965 800	低能量放电
252	0.925 844	高能量放电

根据以上结果, 将该例故障判断为低能量放电兼过热, 与三比值法相比较, 该方法的诊断结果与实际故障更为接近。

例 3: 1990 年重庆凉亭 2 号变压器发生故障, DGA 数据如下: H_2 为 740, CH_4 为 1856, C_2H_6 为 706, C_2H_4 为 3528, C_2H_2 为 78.2 (单位: $\mu L/L$)。三比值编码: 022, 判断为高温过热 (>700 °C)。吊芯结果: 铁心多点接地, 引起过热。该方法诊断结果见表 3。

表 3 聚类分析结果

类中样本	模糊贴近度	样本实际故障原因
70	0.970 646	高温过热
73	0.931 803	螺帽掉在铁心与夹件之中, 造成铁心多点接地
75	0.967 573	高温过热
84	0.988 275	高温过热
102	0.990 715	高温过热
103	0.990 350	高温过热
108	0.987 256	上铁心接地钢带对铁心绝缘空间距离不够, 铁心多点接地
109	0.988 645	A 相铁心多点接地

此故障在该方法中被判断为高温过热, 与三比值法和实际吊芯结果比较, 该方法的判断是正确的, 而且在聚合类中有类似的故障。

利用上述方法对 120 组测试样本进行了故障诊断, 诊断结果与实际故障对比, 无一例明显误判, 判断准确率为 92%。

5 结语

(1) 针对传统模糊 ISODATA 算法的不足进行了改进, 并将改进后的算法应用到变压器故障诊断中, 有效地解决了传统模糊 ISODATA 算法在划分故障类型方面存在的不足。但该算法引入参数较多, 需要在实践中不断优化才能得到理想的效果。

(2) 在故障诊断中, 对于同一种故障类型的变压器 DGA 数据, 由于其原始数据之间的差异, 在聚类中可能未将其分为一类, 而是分为多类, 但这并不影响对新样本的诊断分类。

(3) 在故障诊断中引入模糊贴近度的概念, 更符合人类的思维模式, 为工作人员快速准确地判断故障类型提供了方便, 但同时要求工作人员具备较高的专业素质。

(4) 该算法中指标权向量的确定缺乏理论支持, 在今后的研究中, 宜对这方面做进一步的分析, 以期取得更好的聚类效果。

(下转第 17 页)

反向充电的电流分析与上同。

实际应用中的模拟量隔离调理模块(设定输出范围为 4~20 mA)对输入电信号进行变换的测试数据见表 1。

表 1 对输入电信号进行变换的测试数据

输入电压 /V	全波输入理论输出值/mA	全波输入实际输出值/mA	半波输入理论输出值/mA	正向半波实际输出值/mA	反向半波实际输出值/mA
0	4	3.98	4	3.98	3.98
1	7.2	7.12	6.26	6.23	6.24
2	10.4	10.29	8.53	8.49	8.48
3	13.6	13.42	10.79	10.73	10.76
4	16.8	16.56	13.05	12.99	13.08
5	20	19.96	15.31	15.25	15.33

实验表明,模拟信号隔离调理模块在实际工作时能达到很好的精度,系统实际运行中的情况也证明了这一点。

5 系统抗干扰设计

供电电源:经过 UPS、隔离变压器、滤波器、EMI 抑制器净化,保证系统供电可靠、电压稳定、质量高、抗干扰性好。供电电源与市电之间设有旁路开关,可使它们互相冗余。

控制柜:全部可靠接地,实现全屏蔽。

可编程控制器输出继电器:全部加上阻容吸收回路和压敏电阻。

刀闸监控:全部采用气动。PLC 发出的控制信号控制电控换向阀的动作以开启或者关闭高压气路,利用高压气体的压力开合刀闸主触头来分合主回路;刀闸的辅助触点状态也通过高压气体来驱动气电转换器,从而将刀闸的状态变换为电信号传递给 PLC。

模拟量隔离调理模块:采用前述的数字式串口光纤通信的信号采集与信号回放系统实现。其供电电源采用蓄电池,保证了此模块的工作稳定性以及各控制点与系统间的完全隔离。

通信线路:全部采用光纤连接。一方面是保证控制现场与主控室的绝对隔离以及大量数据的快速传输要求,另一方面则是考虑到在现场强电磁干扰的情况下,能保证从控制柜到主控室这段距离上的可靠通信,并防止试验室由于强放电闪络而影响主控室的正常工作。

软件:设计有纠错处理及陷阱。如在接受上位机控制命令时要求有多重校验密码;采集开关量时要注意防抖判断;程序设计有看门狗模块等。

6 结语

该自动控制系统已经实际投入运行,在调试过程中,系统所采取的隔离措施能够有效地屏蔽现场恶劣环境的影响,保护系统核心硬件的安全;系统设计中的各种优化策略,保证了系统准确高效的数据传输;对模拟量采取的准确有效的监测算法,使控制系统在规定的时限内能做出正确而及时的判断。系统的实际运行表明,在现场高电压、大电流和强电磁干扰的恶劣环境下,系统能够可靠地监测控制电弧充电回路的准确运行,为高压断路器试验的顺利进行提供了前提。同时,它的成功运行对今后的类似工程的建设也具有宝贵的参考价值。

参考文献:

- [1] 王士政. 电网调度自动化与配网自动化技术 [M]. 北京: 中国水利水电出版社, 2003.
- [2] Krishna C M, Kang G Shin. 实时系统[M]. 戴琼海, 译. 北京: 清华大学出版社, 2004.
- [3] Liu Jane W S. Real-time Systems [M]. First Edition, Prentice-Hall, Inc., 2003.
- [4] Oppenheim A V, Willsky A S, Nawab S H. Signals and Systems Second Edition[M]. Prentice-hall, Inc., 1997.
- [5] 阎鸿森, 王新凤, 田惠生. 信号与线性系统 [M]. 西安: 西安交通大学出版社, 1999.

(上接第 13 页)

参考文献:

- [1] 温照森, 胡鸾庆, 邱静. 模式识别与状态监控 [M]. 长沙: 国防科技大学出版社, 1997.
- [2] 张冠军, 钱政, 严璋. 变压器绝缘诊断中的模糊 ISODATA 法[J]. 高电压技术, 1999, 25(1): 1-3.
- [3] 陈守煜. 工程模糊集理论与应用[M]. 北京: 国防工业出版社, 1998.
- [4] 孙才新, 郭俊峰, 廖瑞金, 等. 变压器油中溶解气体分析中的模糊模式多层聚类故障诊断方法的研究 [J]. 中国电机工程学报, 2001, 21(2): 37-41.
- [5] 李金宗. 模式识别导论[M]. 北京: 高等教育出版社, 1994.

欢迎投稿! 欢迎订阅! 欢迎评刊! 欢迎刊登广告!