

变压器油温异常状态识别方法

在变压器本身存在缺陷运行、过负载运行以及冷却器不正常运行时，变压器油温可能会发生异常，影响变压器的安全运行。现有的通过关注变压器油温是否越限的策略，难以在变压器异常运行早期识别异常状态。本文从识别变压器油温异常变化数据出发，采用箱形图和基于近似等频离散化两种分析方法，建立了变压器油温变化异常识别模型，为变压器早期状态预警提供策略。

杜双育 王红斌 李峰/广东电网有限责任公司电力科学研究院

的异常温升过程。因此，如果能够对于变压器的异常温升进行尽早识别，对于提早发现变压器的异常运行状态具有重要的意义。

本文重点从识别变压器油温异常变化数据出发，采用箱形图和基于近似等频离散化两种分析方法，建立了变压器油温变化异常识别模型，实现油温异常数据识别，为变压器早期状态预警提供策略。

1 变压器油温数据异常分析

由于在实际业务中，只关注变压器油温升高的情况，故先将油温数据处理成油温变化率数据，再从变化率数据中筛选出值为正的数据进行研究分析。下面方法介绍中所提到的变化率都为筛选出的值为正的。

定义变化率的指标为：

$$r = \frac{x_{n+1} - x_n}{x_n}$$

式中， x_n ， x_{n+1} 代表本次观测值与下次观测值。

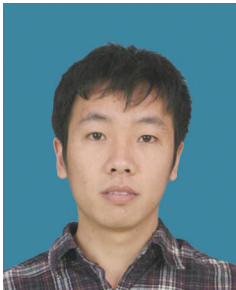
1.1 基于箱形图的变压器油温异常数据分析

相较于 3σ 法则或 z 分数方法是以假定数据服从正态分布为前提的，箱形图的绘制依靠实际数据，不需要事先假定数据服从特定的分布形式，没有对数据作任何限制性要求，它只是真实直观地表现数据形状的本来面貌；另一方面，箱形图判断异常值的标准以四分位数和四分位距为基础，四分位数具有一定的耐抗性，多达25%的数据可以变得任意远而不会较大地扰动四分位数，故可根据箱形图的四分位和上界可将变压器油温的变化率进行分析并划分成几个运行

0 引言

油浸式电力变压器油温是影响变压器运行寿命和负载能力的重要因素，是衡量变压器运行状态的重要特征量之一。变压器本身存在缺陷运行、过负载运行以及冷却器不正常运行时，其油温可能会发生异常，影响变压器的安全运行。但是，我国变压器的温升标准，均以环境温度 40°C 为准，故变压器顶层油温一般不得超过 $40^{\circ}\text{C} + 55^{\circ}\text{C} = 95^{\circ}\text{C}$ 。顶层油温如超过 95°C ，其内部线圈的温度就要超过线圈绝缘物的耐热强度，为了使绝缘不致过快老化，所以规定变压器顶层油温的监视应控制在 85°C 以下。

但是，这种以 85°C 作为油温监视点的运行维护策略，更多关注的是变压器异常运行状态的最终表现，即变压器过负载运行、内部缺陷运行积累到一定程度的表现，而在较大程度上忽略了油温



杜双育/工程师

关键词/Keywords

变压器油温 ·
箱形图 ·
近似等频离散化 ·
预警 ·

区间。

如图1所示,油温变化率的下界非常接近0,故0到下分位可归为第一个区间,该区间油温变化率数据占总数的1/4,变化率都较低,可视为变压器日常运行的正常区间;而下分位到上分位可归为第二个区间,该区间油温变化率数据占总数的1/2,变化率都稍微有点高,视为变压器日常运行的较正常区间;而上分位到上界可归为第三个区间,该区间油温变化率数据占总数比不到1/4,变化率都偏高,视为变压器日常运行的较不正常区间;而上界以上的油温变化率数据较少,变化率都较高,视为变压器日常运行的不正常区间。将正常区间、较正常区间、较不正常区间、不正常区间分别标记为0、1、2、3。

即:

- 0= { r_i | $r_i \in (0, \text{下分数位})$ }
- 1= { r_i | $r_i \in (\text{下分数位}, \text{上分数位})$ }
- 2= { r_i | $r_i \in (\text{上分数位}, \text{上界})$ }
- 3= { r_i | $r_i \in (\text{上界}, +\infty)$ }

式中, r_i 为某变压器的油温变化率。

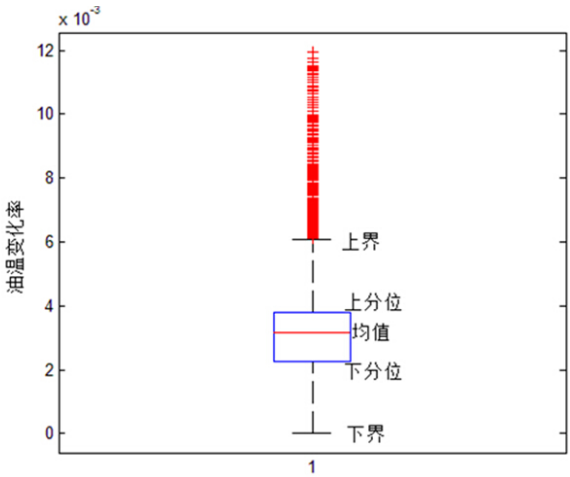


图1 油温变化率箱形

1.2 基于近似等频离散化的变压器油温异常数据分析

近似等频离散化算法(AEFD)是基于数据近似服从正态分布的假设,对连续属性进行离散化。若一个变量服从正态分布,则其观测值落在一个区间的频率与变量在一个区间取值的概率应该相同,利用正态分布变量的分位点将取值区间划分为若干区间,使每个区间的取值概率相同,

进而得到离散区间。

假设将油温变化率取值区间划分为 k 个区间: $(b_i, b_{i+1}] (i=0, 1, \dots, k-1)$, 这里 $b_0 = -\infty$, $b_k = +\infty$, 使正态分布在每个区间取值的概率均为 $1/k$, 离散化后, 以一个符号表示区间 $(b_i, b_{i+1}] (i=0, 1, \dots, k-1)$ 中的每一个油温变化率 r 。

近似等频离散化算法具体有三个步骤如下:

步骤1: 计算划分点, 确定初始划分区间。

按照 $b_i = \bar{r} + Z_{\alpha_i} \cdot \sigma (i=1, 2, \dots, k-1)$ 计算划分点, 这里 $\bar{r}$ 、 σ 分别为属性值的平均值和标准差, Z_{α_i} 为标准正态分布 $\xi \sim N(0, 1)$ 的分位点

$$P(\xi \leq Z_{\alpha_i}) = \alpha_i = \frac{i}{k} (i=1, 2, \dots, k-1)。$$

利用得到的划分点, 得到初始的 k 个划分区间。

步骤2: 合并划分区间, 将包含记录频率很低的区间合并到最接近的区间(尽可能合并到总区间数不超过4个)。

统计各个区间 $(b_i, b_{i+1}] (i=0, 1, \dots, k-1)$ 包含的记录频数及包含记录的均值, 从右往左搜索, 当区间 $(b_i, b_{i+1}] (i=0, 1, \dots, k-1)$ 包含记录频率小于 $1/3k$, 即不到应有频率的 $1/3$ 时, 将该区间合并到最接近的区间, 并修改相应区间的记录频数及包含记录的均值。这个过程重复到没有区间合并为止, 最后得到划分区间。

步骤3: 将连续属性值根据划分区间转换成离散值, 如0, 1, 2, 3, ...。

近似等频离散化算法是以正态分布理论为基础, 根据现实数据在大样本情况近似服从正态分布的特性, 对数据进行离散化。对比朴素贝叶斯离散化分类算法和基于混合概率模型的无监督离散化算法, 该算法思路朴素, 没有复杂的原理, 容易理解, 实现简单。不需要进行大量的分析和计算, 具有线性时间复杂度, 且离散化性能相较于其他算法也是很有效^[1]。

1.3 变压器油温异常数据识别模型

由第1.1和1.2节可知, 箱形图和近似等频离散化方法分别在处理不基于假设数据服从正态分布和基于假设数据近似服从正态分布的情况的优越性, 故综合这两种方法, 分析离散化变压器油温正常运行区间是基于数据分布的较全面考虑的

必要性，以及为结果的准确性。

因此，本文通过将近似等频法离散化的区间置于横坐标轴，将箱形图法离散化的区间置于纵坐标轴，将两种方法所离散化得到的区间对应的离散值相加，建立了变压器油温异常状态矩阵，如图2所示。

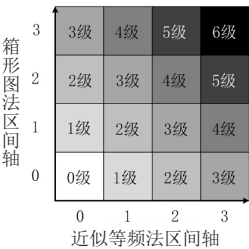


图2 变压器油温异常状态矩阵

图2中，通过对基于箱形图和近似等频离散化方法得到的油温异常数据，经过二维矩阵进行叠加，实现异常数据的综合判据，即为变压器油温所预警的级别。如0级为正常运行，可不采取任何处理措施；1级为较正常运行，需提防该区间的变压器；而2~6级为变压器运行不正常的程度递增级别，需分别采取相应措施处理。

而为了使管理与检修安排不繁杂，将0级、1级区间合并视为正常运行区间，不做任何报警；

将2级区间视为一级报警区间；3级区间视为二级报警区间；4级、5级区间合并视为三级报警区间；6级区间视为严重报警区间。

2 实验结果和分析

随机抽取某电网9台变压器2014年一年的油温数据进行分析，在此对其编号分别为1~9。变压器1部分数据见表2，其中油温数据采样间隔为3 min。

表2 某台变压器部分油温数据

变压器编号	时间	油温	变化率
变压器1	2014-05-12 17:15	45.637 402	0.002 629 4
变压器1	2014-05-12 17:18	45.757 401	0.002 622 483
变压器1	2014-05-12 17:21	45.877 399	0.002 323 628
变压器1	2014-05-12 17:24	45.984 001	5.066 98E-06
变压器1	2014-05-12 17:27	45.984 234	0.002 604 501
变压器1	2014-05-12 17:30	46.104 000	4.836 89E-05
变压器1	2014-05-12 17:33	46.106 230	0.002 554 297
变压器1	2014-05-12 17:36	46.223 999	0.002 629 4
...	...	...	...

根据第2.3节的异常状态矩阵，对9台主变数据分别进行箱形图法统计、近似等频离散化法统计，见表3和表4。

表3 按箱形图法统计结果

主变编号	有效数据	上分位数	超过上分位数的数据个数	上界	超过盒图最大值的的数据个数	数据最大值
1	51 465	0.002 971 732	12 805	0.005 331 115	818	0.005 966 272
2	49 088	0.002 844 680	12 192	0.005 089 382	650	0.005 521 127
3	55 155	0.003 486 216	13 750	0.006 571 864	882	0.007 292 367
4	64 819	0.004 995 837	16 197	0.009 520 183	1318	0.010 588 235
5	64 298	0.004 987 531	16 072	0.009 431 345	1184	0.010 435 301
6	66 564	0.004 900 696	16 638	0.009 384 164	1219	0.010 371 651
7	45 335	0.003 578 399	11 313	0.005 643 006	820	0.006 040 637
8	44 140	0.003 510 312	11 024	0.005 351 409	822	0.005 693 371
9	46 104	0.004 339 740	11 459	0.006 916 996	579	0.007 387 247

表4 按近似等频离散化法统计结果

主变编号	有效数据	90%分位数	超过90%分位数的数据个数	95%分位数	超过95%分位数的数据个数	99%分位数	超过99%分位数的数据个数
1	51 465	0.003 927	5 120	0.004 690	2 612	0.005 643	493
2	49 088	0.003 787	4 883	0.004 418	2 442	0.005 177	490
3	55 155	0.004 830	5 511	0.005 656	2 727	0.006 762	553
4	64 819	0.007 070	6 482	0.008 336	3 243	0.010 022	648
5	64 298	0.006 975	6 434	0.008 207	3 215	0.009 875	643
6	66 564	0.006 901	6 656	0.008 159	3 332	0.009 789	666
7	45 335	0.004 665	4 516	0.005 053	2 257	0.005 805	410
8	44 140	0.004 157	4 562	0.004 897	2 220	0.005 503	424
9	46 104	0.005 569	4 566	0.006 098	2 287	0.007 042	473

综合表3和表4，采用第2.3节所示状态异常状态矩阵，对9台主变数据进行预警，见表5。其中，以1号主变压气为例，随机抽取某一严重告警数据附近数据绘制曲线，如图3所示。

表5 报警情况统计表

主变压气 编号	数据个数	一级报警		二级报警		三级报警		严重报警	
		个数	百分比	个数	百分比	个数	百分比	个数	百分比
1	51 465	7 685	14.93	2 508	4.87	1 761	3.42	325	0.63
2	49 088	7 309	14.89	2 441	4.97	1 785	3.64	160	0.33
3	55 155	8 239	14.94	2 784	5.05	1 831	3.32	329	0.60
4	64 819	9 715	14.99	3 239	5.00	1 924	2.97	670	1.03
5	64 298	9 638	14.99	3 219	5.01	2 030	3.16	541	0.84
6	66 564	9 982	15.00	3 324	4.99	2 112	3.17	553	0.83
7	45 335	6 797	14.99	2 259	4.98	1436	3.17	410	0.90
8	44 140	6 462	14.64	2 342	5.31	1 344	3.04	398	0.90
9	46 104	6 893	14.95	2 279	4.94	1 686	3.66	106	0.23

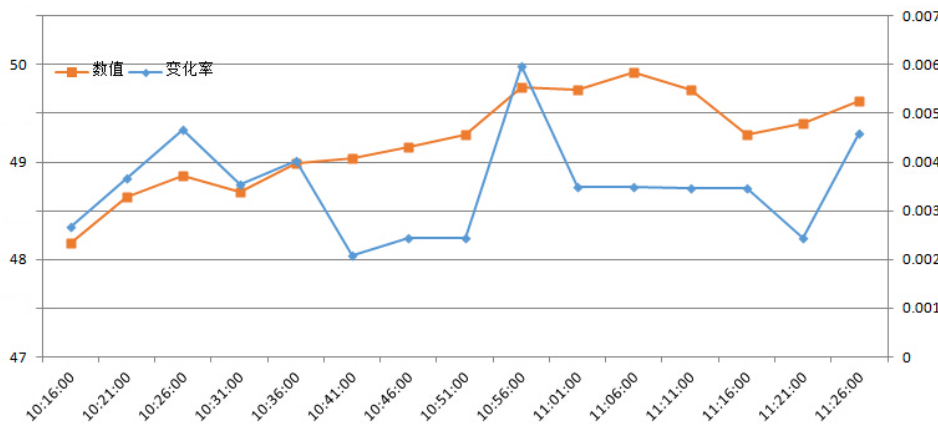


图3 某主变压气异常油温数据情况

可以看出，通过上述识别模型，能够准确定位油温异常突变点，实现异常状态的早期识别。

3 结束语

本文综合了箱形图和近似等频离散化方法的优点，既不局限数据的分布，又考虑了海量数据在概率分析中所满足的大数定律与中心极限定理。为满足油温数据异常识别所提出的解决方案。此外，根据本文所建立的异常状态矩阵，可以实现异常状态的分级告警，为制定差异化的预警策略及处置措施提供支撑。

参考文献

[1] 蒋盛益,李霞,郑琪.一种近似等频离散化方法[J].暨南大学学报.2009,1(30):31-34.
[2] 徐峻岭.基于互信息的无监督特征选择[J].计算机研究与发展.2012,49(2):372-382.

[3] 李刚.基于混合概率模型的无监督离散化算法[J].2002,25(2):158-164.
[4] 徐盈盈.基于集成学习的无监督离散化算法[J].计算机应用.2014,34(8): 2184 – 2187.
[5] 李晓宏,孙林岩,李刚.基于正态分布特征的连续属性无监督离散化方法研究[J].科学与管理.2009,29(6):5-8.
[6] 张文字.数据挖掘过程中连续属性离散化新方法研究[J].数学的实践与认识.2007,37(10):90-95.
[7] 花海洋.一种新的无监督连续属性离散化方法[J].计算机工程与应用.2011,47(6):208-211.

作者简介

杜双育（1985-），男，工程师，工学博士主要从事电力设备状态监测及故障诊断工作。
(收稿日期：2015-12-10) **EA**
(编辑：朱历)