# 基于概率统计和韦布尔分布的一体化终端故障诊断模型研究

**目前各地市供电局存在故障的终端数量比较多，依靠人工排查，工作量比较大，而且对供电局的技术能力要求也比较高。提出了一种基于概率统计和韦布尔分布的终端故障诊断的方法，建立出终端故障诊断的韦布尔模型，通过在线监测终端运行状态和运行数据，实现对故障终端的智能甄别和定位，最大限度减少运维人员的工作量。**

胡姗姗 林伟斌 赵云 张乐平 赖宇阳/南方电网科学研究院

胡珊珊/工程师

*关键词/Keywords*

*一体化终端·*
*故障诊断·*
*概率统计·*
*韦布尔分布·*

## 0 引言

一体化终端及一体化采集与信息集成，是实现对配变的实时数据及低压客户的用电数据进行采集、存储和应用，并按信息化管理的要求，向其他应用子系统提供生产现场的各种基础数据，实现了配电网信息集中管理、统一应用，提高系统集成化程度及降低系统运行维护工作量。

通过在线监测一体化终端运行状态和运行数据归纳出诊断模型，实现对故障终端的智能甄别，以减少运维人员工作量，同时为新入网终端提供运行状况监控，作为入网根据，减少故障来源。

## 1 一体化终端故障诊断模型的建立

### 1.1 终端故障诊断建模

一体化终端常见故障主要有主板故障、电源模块故障、显示模块故障、通信模块故障、控制模块故障、通信故障、485通信故障和接线错误等，本模块通过构建终端故障诊断模型，实现对终端常见故障的自动识别。

终端故障诊断建模主要通过分析电力系统积累下来的大量终端故障历史数据，对各类故障状况分类整理，并给出相应的判定结果，形成专家样本库，最后利用数据挖掘技术将这些数据中蕴藏着的许多潜在的重要因素、事实和关联等有价值的信息提炼出来，分析故障事件同其他因素（如负荷、报警等）的相关性，并进而构建智能诊断模型，最后可以实时处理采集到的用电数据，真正做到高效、实时地进行终端故障诊断。

### 1.2 一体化终端故障诊断模型的具体算法

#### 1.2.1一体化终端故障诊断模型的概率统计算法

终端报警在第$i$天发生报警的次数如表1所示。

表1 终端报警记录表

| 终端报警类别 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 第$i$天的报警次数 | $a_{i0}$ | $a_{i1}$ | $a_{i2}$ | $a_{i3}$ | $a_{i4}$ | $a_{i5}$ | $a_{i6}$ | $a_{i7}$ | $a_{i8}$ | $a_{i9}$ |

终端报警在一定程度能反映终端故障的存在，但由于终端报警存在误报和漏报，而终端出现故障时一般会出现相应的报警，但出现某类报警并不代表故障一定存在。

现将终端报警每天发生的次数转一化，此处提出了类归一化思想的转一化，方便大数据的统计与分析。

定义：通过适当数学工具将非零的数据转化为1，称为转一化。

取

$$\lim_{k\to\infty} b_{ij}^{k} = a_{ij} \qquad (1)$$



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

令

$$w_{ij}=\frac{a_{ij}}{\lim\limits_{k\to\infty} b^{k}_{ij}} \tag{2}$$

即

$$w_{ij}=\begin{cases}1, & a_{ij}>0,\\ 0, & a_{ij}=0,\end{cases} \tag{3}$$

$w_{ij}$为$a_{ij}$的转一元，也是第$i$天，$A_j$故障的开关。当$w_{ij}=1$时，则在第$i$天$A_j$故障开关打开，出现故障$A_j$，当$w_{ij}=0$时，则在第$i$天$A_j$故障开关关闭，不出现故障$A_j$。称$w_{ij}$为故障$A_j$第$i$天的出现元。

假设收集的$n$天数据是足够多的，现统计天里故障出现的次数。

$A_j$故障在$n$天里出现的次数为

$$r_j=\sum_{i=1}^{n} w_{ij} \tag{4}$$

$A_j$故障在$n$天里出现的次数如表2所示。

表2 $A_j$故障在$n$天里出现次数记录表

| 故障类型 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 故障出现元 | $W_{i0}$ | $W_{i1}$ | $W_{i2}$ | $W_{i3}$ | $W_{i4}$ | $W_{i5}$ | $W_{i6}$ | $W_{i7}$ | $W_{i8}$ | $W_{i9}$ |
| 故障出现次数 | $r_0$ | $r_1$ | $r_2$ | $r_3$ | $r_4$ | $r_5$ | $r_6$ | $r_7$ | $r_8$ | $r_9$ |

则故障$A_j$出现率$x_j$为

$$x_j=\frac{r_j}{9n}=\frac{\sum_{i=1}^{n} w_{ij}}{9n}=\sum_{i=1}^{n}\frac{a_{ij}}{9n\lim\limits_{k\to\infty} b^{k}_{ij}}, \tag{5}$$

式中，$i=1,2,\cdots,n$；$j=0,1,2,\cdots,9$。

根据试验数据挖掘得出故障$A_6$与其他9个故障满足以下关系式，即

$$\{A_6\cup(\overline{A_0}\cap\overline{A_1}\cap\cdots\cap\overline{A_5}\cap\overline{A_7}\cap\overline{A_8}\cap\overline{A_9})\}=\{\text{终端故障判断为0}\} \tag{6}$$

而除故障$A_6$外的其他9个故障之间满足以下关系，即

$$\{A_6\cup A_1\cup\cdots\cup A_5\cup A_7\cup A_8\cup A_9\}=\{\text{终端故障判断为1}\} \tag{7}$$

故由串并联关系，该10个故障间的串并联关系图如1所示（10个故障分别为10个不同的开关）。

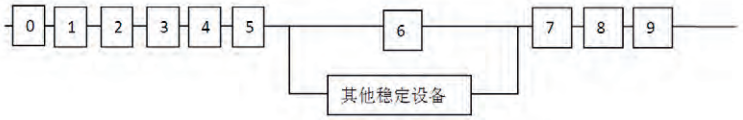


图1 10个故障间的串并联关系图

### 1.2.2韦布尔分布的物理模型

以上的串并联结构是由9个故障所属元件串联而成的（由以上分析可先忽略故障$A_6$），于是可以形象地将结构看成是由9个开关构成的线路，其通电与否取决于最薄弱的开关。单个线路的强度（此处强度为出现率）为一随机变量，设各开关强度相互独立，分布相同，则求线路强度的概率分布就变成求极小值分布问题，由此给出韦布尔分布函数。

以$\zeta$表示单个被断路的出现率，则$\zeta$是一个随机变量。只要一个开关断开，则整个线路断路。只要被一个出现率的故障断开，则整个线路断开，一体化终端故障判断为1，故一体化终端故障判断为1的概率为$p\{\zeta<x\}$，现求$p\{\zeta<x\}$的表达式。注意到单个开关不断开的概率是$1-p\{\zeta<x\}$，令

$$1-p\{\zeta<x\}=e^{-g(x)} \tag{8}$$

若开关的最小出现率为$v$（常数），则出现率小于$v$时线路不断（理想化）。故

$$g(x)\begin{cases}=0, & \text{当}x\leqslant v\\ \geqslant 0, & \text{当}x>v\end{cases} \tag{9}$$

考虑到正幂函数是一种最基本的函数，并且满足$x$愈大时，$g(x)$愈大（出现率愈大，线路断开的概率愈大）的要求，故不妨取

$$g(x)=\begin{cases}\dfrac{(x-v)^m}{x_0}, & \text{当}x>v\\ 0, & \text{当}x\leqslant v\end{cases} \tag{10}$$

式中，$m$、$x_0$为常数，且$x_0>0$，$m>0$，于是得

$$p\{\zeta<x\}=1-e^{-g(x)}=\begin{cases}1-e^{-\frac{(x-v)^m}{x_0}}, & \text{当}x>v\\ 0, & \text{当}x\leqslant v\end{cases} \tag{11}$$

若记$F(x)=p\{\zeta<x\}$，则$F(x)=F'(x)$，即

$$F(x)=F'(x)=\begin{cases}\dfrac{m}{x_0}(x-v)^{m-1}e^{-\frac{(x-v)^m}{x_0}}, & \text{当}x>v\\ 0, & \text{当}x\leqslant v\end{cases} \tag{12}$$

由微积分学中的牛顿－莱布尼茨公式可知

$$p\{a<\zeta<b\}=F(b)-F(a)=\int_b^a F'(x)\mathrm{d}x=\int_b^a f'(x)\mathrm{d}x \tag{13}$$

则可知，单个开关的强度$\xi$所服从的式（12）。故由式（12），可以方便快捷的解决终端故障诊断，不但可以诊断出终端是否有故障，



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

并且能给出哪个部件出现故障的概率最大，且能较准确地知道出现效率的概率。

### 1.3一体化终端故障诊断模型的具体操作过程

对入网终端210天的数据进行监测，如表3所示。

表3 终端报警及故障

| 显示模块故障$A_2$ | 上行通信模块故障$A_3$ | 下行通信模块$A_4$ | 通道故障$A_5$ | 信号不稳定$A_6$ | 485通信故障$A_7$ | 终端抄不到表$A_8$ | 接线错误$A_9$ | 是否故障$Y$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 2 | 0 | 5 | 1 | 0 | 6 | 1 |
| 0 | 1 | 2 | 0 | 12 | 4 | 0 | 2 | 1 |
| 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 1 |
| 2 | 0 | 2 | 0 | 3 | 3 | 0 | 1 | 1 |
| 4 | 0 | 6 | 0 | 0 | 4 | 1 | 0 | 1 |

将表中各故障的出现次数转一化，如表4所示。

表4 终端报警及故障数据预处理

| $W_0$ | $W_1$ | $W_2$ | $W_3$ | $W_4$ | $W_5$ | $W_6$ | $W_7$ | $W_8$ | $W_9$ | $Y$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |

假设收集的210天数据是足够多的。现统计$A_j$故障在210天里出现的次数，即

$$A_j=\sum_{i=1}^{209}w_{ij} \tag{14}$$

统计结果如表5所示。

表5 210天试验数据的统计结果

| 故障类型 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 故障出现元 | $W_{i0}$ | $W_{i1}$ | $W_{i2}$ | $W_{i3}$ | $W_{i4}$ | $W_{i5}$ | $W_{i6}$ | $W_{i7}$ | $W_{i8}$ | $W_{i9}$ |
| 故障出现次数 | 32 | 9 | 17 | 14 | 14 | 12 | 89 | 14 | 25 | 17 |

则故障$A_j$出现率$x_j$为

$$x_j=\frac{r_j}{9n}=\frac{\sum_{i=1}^{n}w_{ij}}{9n}=\sum_{i=1}^{n}\frac{a_{ij}}{9n\lim\limits_{k\to\infty}b_{ij}^{k}}, \tag{15}$$

式中，$i=1, 2, \cdots, n$；$j=0, 1, 2, \cdots, 9$。

终端故障类型出现率如表6所示。

表6 终端故障类型出现率

| 故障类型 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 故障出现次数 | 32 | 9 | 17 | 14 | 14 | 12 | 89 | 14 | 25 | 17 |
| 出现率 | 0.016 93 | 0.004 76 | 0.009 | 0.007 41 | 0.007 41 | 0.006 35 | 0.047 1 | 0.007 41 | 0.013 23 | 0.009 |

除故障$A_6$以外的9个故障中，出现率最小的是故障$A_1$，其出现率为0.004 76，故$\nu=0.004\ 76$，而由于数据太小，结果误差会很大，故此允许对所有相应数据增大到1 000倍，得到相应的$x'$。而相应增大1 000倍的$\nu$重记为$\nu'$。根据实际情况，考虑形状参数$m=\frac{1}{2}$，尺度参数$x_0=\frac{a_j}{2}$。

得到的韦布尔密度函数为

$$f(x)=F'(x)=\begin{cases}\dfrac{m}{x_0}(x-v)^{m-1}e^{-\frac{(x-v)^m}{x_0}}, & x>v\\ 0, & x\le v\end{cases}$$

可得10种故障终端诊断的实际模型，即韦布尔模型为

$$f(x')=F'(x')=\begin{cases}\dfrac{1}{a_j}e^{-(x'-v')^{\frac{1}{2}}}, & x'>v'\\ 0, & x'\le v'\end{cases}$$



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

由建立的韦布尔模型可以得出，当终端故障判断为1时，各故障出现的概率如表7所示。

表7 终端诊断概率结果

| 故障类型 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 出现率 | 0.016 93 | 0.004 76 | 0.009 | 0.007 41 | 0.007 41 | 0.006 35 | 0.047 1 | 0.007 41 | 0.0132 3 | 0.009 |
| 终端诊断概率 | 0.935 4 | 0.111 1 | 0.461 1 | 0.363 8 | 0.363 8 | 0.294 1 | 0.047 1 | 0.363 8 | 0.734 5 | 0.461 1 |

故在期间出现的漏报，由$A_0$引起终端故障的可能性为93.54%，引起终端故障的可能性为73.45%，而$A_1$和$A_7$引起终端故障的可能性分别只有11.11%和4.71%。

当出现如表8的数据时，确定漏报的概率$P\{A_0 \cup A_1 \cup \cdots \cup A_9\} = \sum_{i=0}^{9} = x_i = 0.128\ 6$。

表8 漏报类型数据

| 终端报警类别 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 报警次数 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

当出现类似表9的数据时，确定误报的概率 $P\{A_0 \cap A_1 \cap \cdots \cap A_9\} = \Pi^{9}_{i=0}(1 - x_i) = 0.877\ 9$。

表9 误报类型数据

| 终端报警类别 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 报警次数 | 0 | 11 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 |

## 2 模型验证

在210天试验中，有7天的数据如表8所示，在解决故障后记录了故障的出处并与模型结果进行对比，如表10所示。

表10 模型验证结果表

| 故障类型 | $A_0$ | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ | $A_7$ | $A_8$ | $A_9$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 故障出现次数 | 7 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 5 | 0 |
| 概率 | 1 | 0 | 0 | 0 | 0 | 0.285 7 | 0 | 0 | 0.714 3 | 0 |
| 原诊断概率 | 0.935 4 | 0 | 0.001 2 | 0.003 1 | 0 | 0.294 1 | 0 | 0 | 0.734 5 | 0.000 6 |
| 绝对误差 | 0.064 6 | 0 | 0.001 2 | 0.003 1 | 0 | 0.008 4 | 0 | 0 | 0.020 2 | 0.000 6 |

如此可见，所构建的韦布尔模型是合理的。且试验数据越多，模型构建将更为精准。

## 3 结束语

提出基于概率统计和韦布尔分布的终端故障诊断。通过这两种方法，建立终端故障诊断的韦布尔模型，实现了直接判断出终端是否故障，从而节约了维护终端的工作量。所提出的方法不但可以诊断出终端是否有故障，并且能给出哪个部件的故障概率最大，且能较准确地知道故障出现的概率，最大限度地减少运维人员的工作量。

### 参考文献

[1] 王妍. 概率统计在实际问题中的应用举例[J]. 中国传媒大学学报，2007，14(1)：15-19.

[2] 翁汉琍，毛鹏，林湘宁. 一种改进的电网故障诊断优化模型[J].电力系统自动化，2007，31(7)：66-70.

[3] 赵伟，白晓民，丁剑，等. 基于协同式专家系统及多智能体技术的电网故障诊断方法[J]. 中国电机工程学报，2006，26(20)：1-8.

[4] 方江林. 韦布尔分布秩序统计量的随机比较[J]. 内江师范学院学报. 2008，23(6)：33-34.

[5] 邓集贤. 概率论及数理统计[M]. 北京：高等教育出版社，2009.

（收稿日期：2015－05－26）EA

作者简介

胡珊珊（1981-），女，工程师，主要研究方向为电能计量。



(C)1994-2023 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net