

家用电器故障实时检测

摘要

家用电器的故障是我们生活中不可完全避免的。随着电器使用时长的增加，电器组成器件的老化会使其工作能力呈现衰减趋势，电器老化的成因有很多种，同时也是不可避免的。为了保证电器器件的工作性能，就需要对电器的组成元件进行实时监测。这样才能在电器器件出现故障时即时停止使用，防止故障程度的升级，同时尽快维修电器，尽快恢复使用。因此，正确且即时地检测出故障将对家用电器的维修费用和使用寿命有极大的改善。

针对问题一，传感器读取到异常数据即会给使用者带来误判信息，可能出现将故障判断为正常的情况，此时对于异常数据的处理方式是解决此问题的关键点。根据肖维勒准则剔除异常数据，小波去噪分析减少数据波动两个原理，我们将其引入问题一，并运用 MATLAB 软件编写相应程序，从而建立起解决问题的模型，消除异常数据对故障判别的影响。

针对问题二，我们对附件一 A 样本中正常与故障表格的数值型数据筛选出来，再由肖维勒准则和小波分析的 MATLAB 程序进行处理，然后把这些数据文件分别导入到 spss 软件中。利用 spss 里的系统聚类进行分析，提取出 5 个主要参数。再由比较均值，将两种情况下的 5 个主要参数均值求出，接着引入故障判别半径这一概念，对接下来的问题进行求解。

针对问题三，我们通过上述数据的处理方式及过程，将附件二中的 C 组数据进行处理，然后由故障判别半径公式得出： $r_C > r_B$ 这一结果，得知上述问题所建立的模型依然适用附件二。在判断附件三中的数据时，我们通过观察数据分布特点将数据划分为两部分，然后由同上方式处理数据，得知 D 为故障，E 为正常。

针对问题四，在问题二中我们对故障判别半径进行了调整，由此可看出模型的故障判别半径一直在优化，因此在假定故障发生随机的情况下，模型会出现误判。接着，我们分析了将正常判断为故障和将故障判断为正常这两种错误造成的成本，得出将故障判断为正常所付出的成本更大，更应该避免。于是，我们为了优化故障判别半径引入了半径限定参数 λ 将故障判别半径限定在一个更小的范围内，使故障判别半径能够更加精准地检测电器的实时状态。

针对问题五，考虑到实际生活中，家用电器的控制器计算能力有限，因此我们在保证准确性的前提下建立了人工神经网络模型以降低计算复杂度。首先，我们对数据进行了 0-1 标准化，降低了数据的数量级；然后利用人工神经网络来预测家电正常工作时的数据，再把实测数据与预测数据进行简单的比较来检测家电的工作状态。从计算过程和计算次数上可看出，此问题建立的模型极大地降低了计算的复杂度。

关键词：肖维勒准则 小波去噪分析 故障判别半径 半径限定参数 人工神经网络

1 问题重述

家用电器在日常生活中必不可少，电器老化同时也是不可避免的，当用户提出电器需要维修时，往往电器的故障情况已经十分严重。为了保证用户的体验效果，现需要一种能够对电器运行状况进行实时监测并判别的方法来解决该问题。

问题 1: 在复杂工作环境下的电器，有可能会产生导致传感器读取到异常数据，针对此问题给出解决方案。

问题 2: 根据附件一中的数据分析出不同参数之间的相关性及其对故障判别的重要程度。结合之前的工作，建立一个故障判别模型对附件一中的数据进行判别。

问题 3: 判断问题二所建立的模型是否依然使用附件二中的数据，若不适用给出修正方案，再尝试判断附件三中的数据，判断其状态为正常或故障。

问题 4: 考虑模型是否会出现误判，若出现误判，也要说明将正常判断为故障和故障判断为正常这两种错误哪一种更应该避免。并嵌入一个影响因子，给出解决方案。

问题 5: 由于家用电器的控制器的计算能力有限，无法解决计算量特别大的模型。要在保证判断准确的前提下尽可能地降低计算复杂度，并给出解决方案。

2 模型假设

1. 假设剔除的数据均为异常数据。
2. 假设字符型变量对电器故障判别无影响。
3. 假设半径限定参数合理。
4. 数据来源真实可靠。

3 符号说明

v_i	残差
X_i	测量值
\bar{X}	平均值
σ	标准偏差
ω_n	肖维勒系数
$\Psi(t)$	平方可积函数
R	实数集合
$L^2(R)$	实数域平方可积空间
$Wf(a, b)$	小波变换函数
a	伸缩因子
b	平移因子
r	故障判别半径
V_i	参数的均值

λ	半径限定参数
x_{max}	样本数据的最大值
x_{min}	样本数据的最小值
y	节点输出
x_i	节点接受的信息
ω_j	相关连接权重
θ	输入信息的阈值
n	节点数
Z	学习因子
W	误差修正量
T	势态因子

4 模型的建立与求解

4.1 问题一

4.1.1 问题一的分析

电器在复杂的工作环境下工作时，由于外界的温度、压力等对电器内部零件会造成一定的影响，包括传感器。此时，有可能导致传感器读取到异常数据，并由此给人们传达出错误的故障判别分析，而影响人们对电器的使用和实时监测。传感器读取到异常数据时，如果经过系统的数据清理过程，即利用肖维勒准则剔除异常值，对于电器接下来的工作和运转具有很大的帮助和实际的意义。将异常值剔除后，电器继续运作的过程中仍然少不了对故障的监测和分析这一工作，此时，需要对系统传出的大量信号进行小波分析(Wavelet Analysis)，因为故障诊断中的首要问题就是对观测信号的故障特征提取，即对观测信号进行信号处理，从中获取反映故障信息的特征。由于故障诊断中所遇到的信号绝大多数都是非平稳信号，而特别适用于非平稳信号处理的工具就是小波分析。

小波变换是时间(空间)频率的局部化分析，它通过伸缩平移运算对信号(函数)逐步进行多尺度细化，最终达到高频处时间细分，低频处频率细分，能自动适应时频信号分析的要求，从而可聚焦到信号的任意细节，解决了 Fourier 变换的困难问题，成为继 Fourier 变换以来在科学方法上的重大突破。因此小波变换也有了“数学显微镜”的美称。

在此，我们利用肖维勒准则和小波分析来帮助解决复杂工作环境下传感器读取到异常数据的问题，具体模型如下。

4.1.2 问题一模型的建立

(1) 肖维勒准则剔除异常数据模型建立

肖维勒准则认为：当残差

$$|v_i| = |X_i - \bar{X}| > \omega_n \sigma \quad (1)$$

其中, X_i 为测量值, \bar{X} 为平均值, σ 为标准偏差。

测定值 X_i 就是含有粗差的可疑数据, 应当予以剔除。通过对该方法的 n 次测量结果进行分析, 如果某误差出现的次数小于 $1/2n$ 次时, 则将其剔除。按照下列近似公式, 通过 MATLAB 程序求解:

$$\omega_n = 1 + 0.4 \ln(n) \quad (2)$$

其中, ω_n 为肖维勒系数。

以工作环境 A (正常) 中参数 2 为例:

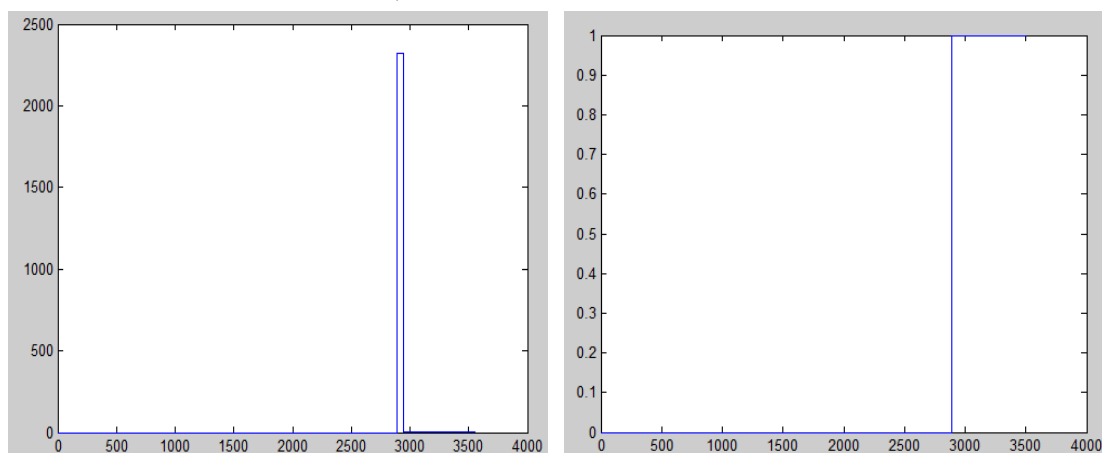


图 4-1 肖维勒准则剔除异常数据前后图示

(2) 小波分析模型建立

小波变换的基本概念

设 $\Psi(t)$ 为一平方可积函数, 即 $\Psi(t) \in L^2(R)$, 若傅里叶变换 $\Psi(\omega)$ 满足条件:

$$C_\Psi = \int_R \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (3)$$

则称 $\Psi(t)$ 为一个基本小波或小波函数, 我们称 (3) 为小波函数的可容性条件, 其中 t 为时间, ω 为频率, R 为实数集合, $L^2(R)$ 为实数域平方可积空间, 由函数 $\Psi(t)$ 经过伸缩和平移得到的一族函数:

$$\Psi_{ab}(t) = \frac{1}{\sqrt{a}} \Psi \left[\frac{t-b}{a} \right] \quad (4)$$

称为小波函数或依赖于 a, b 的连续小波, 式中 a, b 为实数且 $a \geq 0$, a 为伸缩因子, b 为平移因子。任意信号 $f(t) \in L^2(R)$, 其小波变换 $Wf(a, b)$ 定义为:

$$Wf(a, b) = \langle f(t), \Psi_{ab} \rangle = \frac{1}{\sqrt{a}} \int \Psi \left[\frac{t-b}{a} \right] f(t) dt \quad (5)$$

由上式可知 a 的变化不仅改变连续小波的频谱结构, 也改变其窗口的大小与形状。随着 a 的减小, $\Psi_{ab}(t)$ 的频谱就向高频方向移动, 而 $\Psi_{ab}(t)$ 的宽度则越来越狭小。这就满足了信号频率高相应的窗口应该小, 因而它在时间 (或空间) 域上均有较

高的分辨力。

小波变换是可逆的，则信号 $f(t)$ 的重构公式

$$f(t) = \frac{1}{C_h} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Wf(a, b) \Psi_{ab}(t) \frac{1}{a^2} da db \quad (6)$$

式中：

$$C_h = \int_{-\infty}^{+\infty} \frac{|\hat{\phi}(\omega)|^2}{|\omega|} d\omega \quad (7)$$

以工作环境 A（正常）中参数 3 为例：

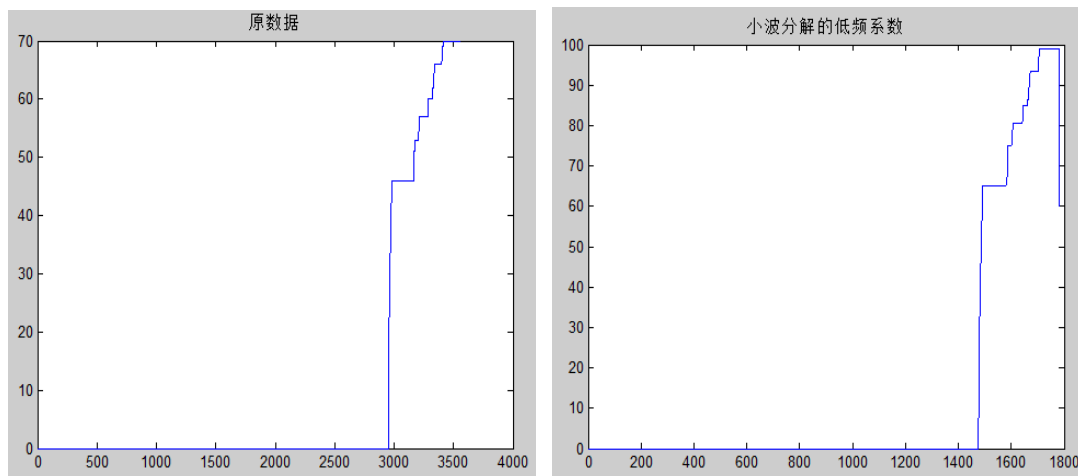


图 4-2 小波去噪前后图示

在数学上，小波去噪问题的本质是一个函数逼近问题，即如何在由小波母函数伸缩和平移版本所展成的函数空间中，根据提出的衡量准则，寻找对原信号的最佳逼近，以完成原信号和噪声信号的区分。小波域阈值去噪，将信号映射到小波域，根据噪声和噪声的小波系数在不同尺度上具有不同的性质和机理，对含噪信号的小波系数进行处理。我们把处理传感器读取到异常信号这一过程模拟为小波域阈值去噪过程，系统根据阈值去噪原理，将传感器读取到的异常信号进行平滑处理，使其“消失”，那么传感器读取到的异常信号会瞬间被阻挡，就不会给使用者带来错误的故障判别。具体处理过程会在 MATLAB 上实现。

4.2 问题二

4.2.1 问题二的分析

在问题一中，我们对于传感器读取到异常数据的情况给出了具体处理措施，所以在问题二里，对于附件一中给出的大量经脱敏处理的数据，我们将工作环境 A 的“正常”与“有故障”两个样本中的数值型数据筛选出来，先分别在肖维勒方法的 MATLAB 程序中剔除样本中的异常数据，然后再分别经过小波分析程序去噪。将处理后的两个数据文件分别导入到 spss 软件中进行下一步的分析。由于所提供的附件一中 A 的参数有 17 个之多，在一定程度上影响我们对参数的重要

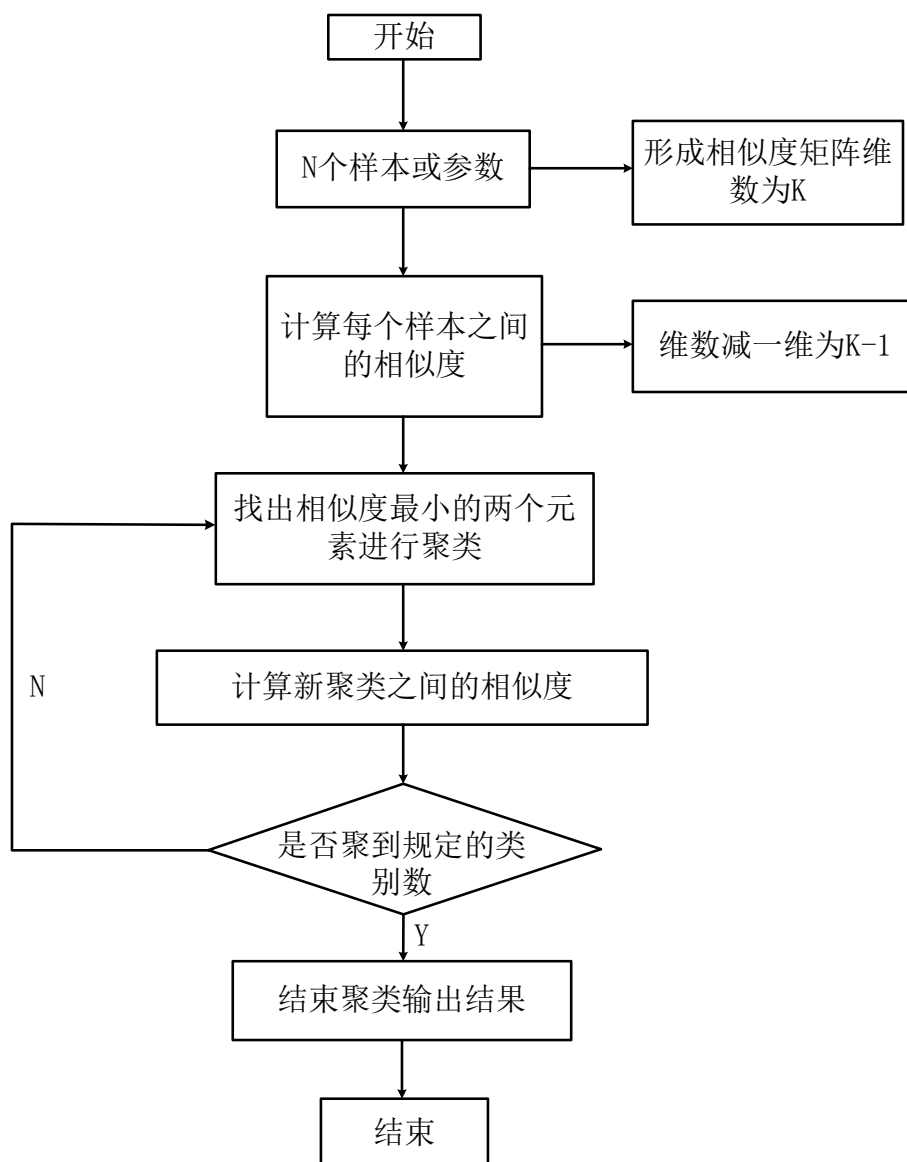
程度进行判别，所以我们利用 spss 里分析中的系统聚类，来对 17 个参数进行聚类合并，并由此提取出 5 个主要参数。然后又利用 spss 里的主成分分析，来检验参数提取的是否正确，能够更加有力地说明提取出的 5 个参数的准确性。

将参数提取之后，返回到 spss 的数据视图中，将其余参数列删除只留下 5 个参数列，再根据分析里的比较均值选项，在输出窗口可以看到 5 个参数的均值。在这里我们引入故障判别半径这一概念，来判别附件一中 B 的电器故障情况。

4.2.2 问题二的数据处理

(1) 提取主要参数

把经过肖维勒方法、小波分析两个程序处理后的正常与有故障情况下的数据文件导入 spss 中，通过聚类分析，能够得到相应结果。实现聚类分析的流程图如下所示：



得到 A 正常情况下的有关图表分析如下：

表4-1 聚类结果显示

聚类表						
阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	1	12	.000	0	0	3
2	13	14	109.260	0	0	3
3	1	13	551.430	1	2	4
4	1	17	1014.125	3	0	5
5	1	10	2955.328	4	0	11
6	7	8	56326.000	0	0	7
7	7	16	152901.000	6	0	9
8	9	11	204729.000	0	0	10
9	7	15	712521.667	7	0	10
10	7	9	1237042.250	9	8	12
11	1	2	1984822.047	5	0	12
12	1	7	3654087.273	11	10	15
13	5	6	25729029.000	0	0	15
14	3	4	52150272.000	0	0	16
15	1	5	251800000	12	13	16
16	1	3	379100000	15	14	0

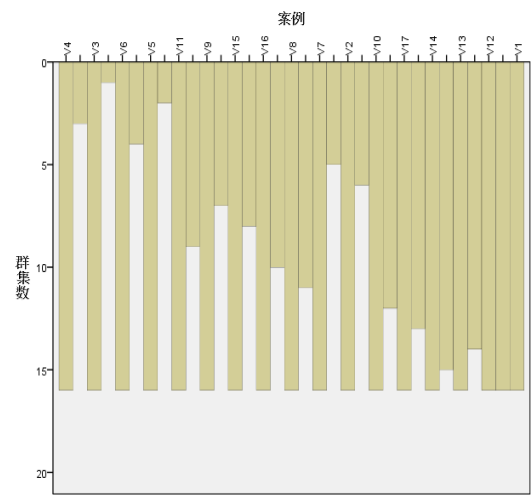


图 4-3 聚类结果冰柱图

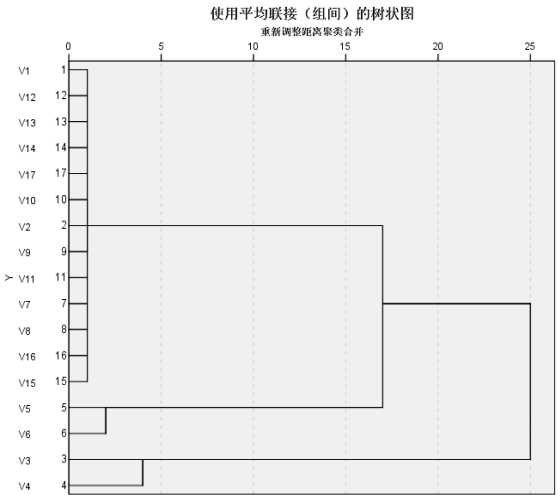


图 4-4 聚类结果树状图

由图我们可以从中提取出主要参数，分别是：V1，V5，V6，V3，V4。
为了验证提取出的参数的准确性，我们采用主成分分析方法对 A 正常情况来做进一步的提取参数工作。
具体分析结果如下所示：

表 4-2 主成分分析结果

成份矩阵^a

	成份			
	1	2	3	4
V2	.971	-.190	-.109	.054
V3	.954	-.163	-.095	.160
V4	.967	-.192	-.115	.029
V5	-.164	.261	-.052	.933
V6	.579	.512	.497	-.134
V7	.493	.338	.609	.044
V8	.917	.137	.258	.057
V9	.883	-.188	-.170	-.155
V10	.964	-.192	-.118	.007
V11	.968	-.003	.098	-.111
V13	.308	.789	-.520	-.091
V14	.308	.789	-.520	-.091
V16	-.941	.126	.204	.109
V17	.721	.364	.470	-.004
V1	.951	-.155	-.096	.170
V12	.951	-.155	-.096	.170

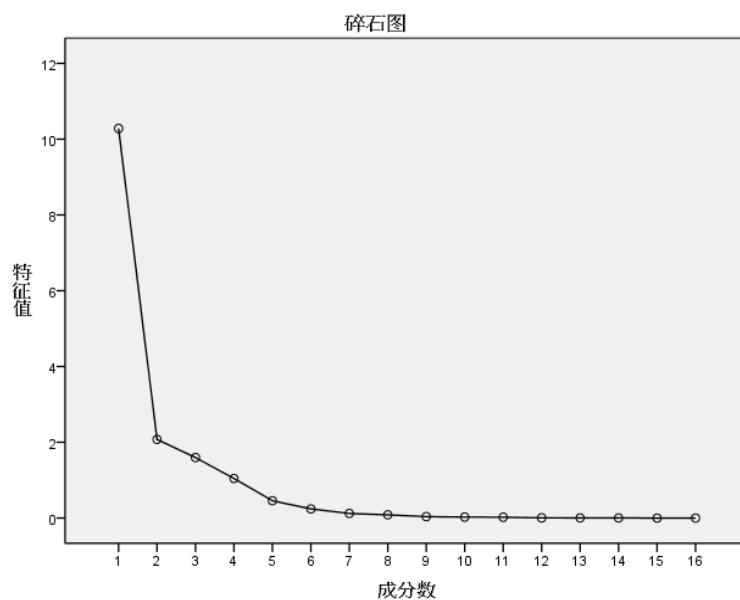


图4-5 主成分分析碎石图

由主成分分析图示结果可知，聚类分析方法提取的 5 个参数在主成分中占有很大比重，能够反映整体信息，达到了验证的目的。

（2）计算主要参数均值

由上述分析结果后，我们将 A 正常情况下的数据视图中保留下V1,V5,V6,V3,V4 参数列数据，将其他数据列删除。然后在分析里的均值选项中，求得 5 个参数的均值，结果如下所示：

表 4-3 A 正常的主要参数均值

	V1	V3	V4	V5	V6
均值	.18	157.14	117.70	249.27	305.74
N	3509	3509	3509	3509	3509
标准差	.381	340.732	268.221	25.647	57.250

表4-4 A故障的主要参数均值

	V1	V3	V4	V5	V6
均值	1.00	575.26	119.10	293.46	309.76
N	3128	3128	3128	3128	3128
标准差	.000	428.541	125.209	48.014	43.348

经过数据视图表头名称和附件中A表中的参数名称比对，得知：V1—参数2，V5—参数6，V6—参数7，V3—参数4，V4—参数5。5个参数均值之间的差值大小，说明了它们在正常和有故障情况下的值的变化大小，进而也说明出它们和故障产生的关联程度，所以也能够由此得出参数对故障判别的重要程度为：

$$\text{参数4} > \text{参数5} > \text{参数6} > \text{参数7} > \text{参数2}$$

4. 2. 3 问题二的模型建立

由5个主要参数的差值数据，我们引出一个故障判别半径概念，即

$$r = \sqrt{\sum_{i=1}^n (V_{i\text{故}} - V_{i\text{正}})^2} \quad (n \text{ 为正整数}) \quad (8)$$

在工作环境 A 情况下，我们把主要参数均值代入公式，如下：

$$\begin{aligned} r_A &= \sqrt{(1 - 0.18)^2 + (575.26 - 157.14)^2 + (119.10 - 117.70)^2 + (293.46 - 249.27)^2 + (309.76 - 305.74)^2} \\ &= 420.4710 \end{aligned}$$

在此结果下，我们暂定 r_A 为故障判别半径标准值，其余工作环境下求得的 r 必须满足： $r > r_A$ ，才可判定电器出了故障；而当 $0 < r < r_A$ 时，那么故障判别半径的标准值会“更新”， r 就是新的半径标准值。

在此模型的基础上，我们对附件一中的 B 工作环境下的数据在 spss 里进行了分

析，得到主要参数的均值如下图所示：

表 4-5 B 正常的主要参数均值

	参数2	参数4	参数5	参数6	参数7
均值	1.22	293.09	64.75	260.11	324.52
N	3516	3516	3516	3516	3516
标准差	39.170	419.260	108.151	62.220	10.313

表 4-6 B 故障的主要参数均值

	参数2	参数4	参数5	参数6	参数7
均值	.52	456.23	279.90	316.47	346.78
N	3516	3516	3516	3516	3516
标准差	.499	449.224	302.596	17.270	12.726

由公式（8）可得：

$$\begin{aligned}
 r_B &= \sqrt{(1.22 - 0.52)^2 + (293.09 - 456.23)^2 + (64.75 - 279.90)^2 + (260.11 - 316.47)^2 + (324.52 - 346.78)^2} \\
 &= 276.7248
 \end{aligned}$$

可知 $r_B < r_A$ ，再结合 B 工作环境的已知情况，可知故障判别半径标准值更新至： $r = 276.7248$ 。

4.3 问题三

4.3.1 问题三的分析

在上述问题二中，我们对于附件一的解决模型已经建立。对于附件二里工作环境为 C 的数据，我们依然使用上述建立的故障判别半径模型来解决此文件中的数据。经过计算半径后，我们发现 $r_C = 325.6786$ ，很明显 $r_C > r_B$ ，则 C 工作环境下的电器故障判别准确。由此，我们可得出附件一中所建立的模型依然适用于附件二中的数据。

得出此结论后，我们直接将附件三中的 D、E 工作环境数据在 MATLAB 程序中进行处理，仔细观察可知，D 环境下的总时长为 31 分 29 秒，E 环境下的总时长为 1 小时。对于 D、E 工作环境下的电器故障判别，只有一个表格可以去分析，和前两个附件中所提供的两个表格对比分析的情况不一样，所以，为了使用已建立的模型来判别故障情况，我们分别将 D、E 中的数据以平分时间段的方法把数据划分为两部分，放在两个表格中。然后在 spss 中，每种环境下都能得出两个均值表来计算故障判别半径值，和标准值进行比对后，得出结论：D 为故障，E 为正常。

4.3.2 问题三的数据处理

(1) C 工作环境下数据处理

依照前边所述，我们分别将附件二中的两个表格中的参数 2, 4, 5, 6, 7 提取出来，删除其余数据，然后利用肖维勒、小波去噪的 MATLAB 程序给数据进行筛选处理。在 spss 软件里打开处理过的数据表格，在分析里的均值选项，给数据作出均值结果，具体如下所示：

表 4-7 C 正常的主要参数均值

	参数2	参数4	参数5	参数6	参数7
均值	.49	401.13	211.47	318.00	347.60
N	3551	3551	3551	3551	3551
标准差	.500	446.664	258.046	14.325	10.541

表 4-8 C 故障的主要参数均值

	参数2	参数4	参数5	参数6	参数7
均值	252.80	255.06	69.90	293.01	327.52
N	3451	3451	3451	3451	3451
标准差	722.933	402.931	122.639	35.968	13.563

根据公式（8），将两个表中各参数的对应均值作出如下计算：

$$r_c = \sqrt{(252.8 - 0.49)^2 + (255.06 - 401.13)^2 + (69.9 - 211.47)^2 + (293.01 - 318)^2 + (327.52 - 347.6)^2} \\ = 325.6786$$

由前边计算的结果，可知 $r_c > r_B$ ，则由故障判别半径定义得，附件一中模型对 C 工作环境故障与正常情况验证正确，依然适用。

(2) 附件三数据处理

依照前边处理附件数据的方式，将附件三中两个表格分别导入 MATLAB 中提取出 5 个主要参数指标数据列。

对于 D 环境：

观察处理后的数据表，可以看出每个参数所对应的列数据在所给时间段内分布较均匀，总行数为 1752，则将其数据根据时间行平均分，取 1-876 行为 D 前半部分，取 877-1752 行为 D 后半部分，分别放在两个表格中。打开 spss，分别打开数据文件 D 前、D 后，由分析中的均值选项，可以得出两个均值表：

表 4-9 D 前半部分主要参数均值

	V1	V2	V3	V4	V5
均值	1.00	707.31	173.82	250.06	330.05
N	875	875	875	875	875
标准差	.000	361.495	116.049	79.023	11.616

表4-10 D后半部分主要参数均值

	V1	V2	V3	V4	V5
均值	1.00	440.57	95.45	254.72	327.61
N	875	875	875	875	875
标准差	.000	446.945	121.398	46.202	9.908

根据公式 (8)，可列出下式：

$$r_D = \sqrt{(707.31 - 440.57)^2 + (173.82 - 95.45)^2 + (250.06 - 254.72)^2 + (330.05 - 327.61)^2} \\ = 278.0643$$

得： $r_D > r_B$ ，则可以判断 D 工作环境下电器有故障。

对于 E 环境：

观察处理后的数据表格，我们分析出在前边一小段时间内，参数 2, 4, 5 的值均为 0，所以只按照整个时间段来平分数据，固然得不到有说服力的结果。而当参数 2 由 0 变为 1 后，参数 4, 5 也开始由 0 发生数值变化，所以我们将参数 2 从 0 变为 1 的时间节点开始，将后边的 1650 行数据平分，1813-2638 行为 E 前半部分，2639 行-3462 行为后半部分，同上所述，将两个表格分别导入 spss 中，求得均值表如下所示：

表 4-11 E 前半部分主要参数均值

	V1	V2	V3	V4	V5
均值	1.00	893.82	565.76	279.81	344.87
N	825	825	825	825	825
标准差	.000	49.738	127.803	46.668	6.298

表 4-12 E 后半部分主要参数均值

	V1	V2	V3	V4	V5
均值	1.00	900.00	640.00	300.00	348.05
N	823	823	823	823	823
标准差	.000	.000	.000	.000	2.866

根据公式 (8)，可列出下式：

$$r_E = \sqrt{(893.82 - 900)^2 + (565.76 - 640)^2 + (279.81 - 300)^2 + (344.87 - 348.05)^2} \\ = 77.2497$$

在对于工作环境已知而电器状况未知的情况下，故障判别半径起到了相当大的作用，由得出数据可知： $r_E < r_B$ ，则在一定程度上可判断出 E 工作环境下电器正常。

4.4 问题四

4.4.1 问题四的分析

由前述问题的解决方案可知，虽然可以在机器工作时进行实时计算从而得出

相应的故障判别半径从而做出初步判定机器是否发生了故障，但是以上所做模型的故障判别半径是根据实验不断要做更新的，也就是说这个半径的值只能通过多次实验多次更新不断接近故障临界值，但很难达到故障临界值，因此就有可能在判定机器的故障半径时使此半径介于故障临界值与当前所更新临界半径间，从而无法确认机器是否故障导致误判。

若出现误判情况，则会出现将正常误判为故障与故障误判为正常两种情况，为了分析哪种情况对于我们现实生活中的影响更大，我们引入了维修成本、机器停运损失等参数，对两种情况作出分析，并以实际例子进行计算，得出结论，从实际生活与所计算的结果上看，故障判定为正常的损失会更大，所以我们应该尽量避免将故障判断为正常这种情况，因为若将故障判断为正常可能会导致机器长时间处于故障工作状态，从而减少机器寿命并损坏机器零件，甚至会造成电路故障导致火灾的发生。正是由于要避免这种情况，上述模型仅能做到避免将正常判断为故障这种情况，所以可加以改进。

为了尽量避免将故障判断为正常这种情况的产生，我们引入半径限定参数 λ ，从而逐步加快缩小故障判定半径，可以使出现故障误判为正常的概率降低，通过多种环境的机器运行便可不断改进更新半径，从而使得误判几率逐渐降低，增加模型的可行性。

4. 4. 2 两种误判情况的计算处理

假设维修时间成本分别为 A_1, A_2 （指顾客为想得到所期望的商品或服务而必须耗费的时间换算而成的代价）、维修财务成本为 B ，总维修成本为 Z 。

表 4-13 维修成本分析表

维修成本 Z	时间成本 A	财务成本 B
正常判断为故障	A_1	无
故障判断为正常	A_2	B

由表可看出正常判断为故障时只有时间成本，而故障判断为正常时时间成本和财务成本都有。同时，结合实际分析的话，故障判断为正常时电器原故障程度可能会升级，进而增加了故障维修的财务成本，正常判断为故障的话，有检查故障的时间成本，故障判断为正常会产生维修时间成本，维修时间成本往往比检查时间成本大，因此 $A_2 > A_1$ ，由总成本 $Z = A + B$ 可知故障判断为正常时所付出的代价或成本更大。因此，结合实际和表 6-1 可判断出故障判断为正常更应该避免。

4. 4. 3 半径限定参数的引入

由于要求尽量避免故障判断为正常的情况，因此我们在模型中引入半径限定参数 λ ，我们使得最近更新半径与半径限定参数的乘积作为最终更新半径，即为

$$R_{new} = R_{update} \times \lambda \quad (9)$$

根据经验判断，我们将 λ 取为 0.9，从而将半径限定在一个更小的范围内，即可减小故障判断为正常的概率。

4. 5 问题五

4. 5. 1 问题五的分析

考虑到家用电器的控制器的计算能力有限，于是我们需要降低数据处理的计算复杂度。而前面的故障判别模型先将主要数据求均值，然后再求故障判别半径，最后将实验组求出来的故障判别半径与求解模型得出的故障判别半径进行比较得出电器是否有故障。由此可见，原有的诊断故障的计算量是十分复杂的。于是，

我们直接使用人工神经网络模型对未来时段的数据（未来时段的预测数据是正常情况下的预测数据）进行预测，然后将实时短时间段内的数据和预测数据进行比较，如果数据值基本吻合那么电器在这一实时阶段的状态是正常的。

4.5.2 数据的预处理

由于不同变量单位不同，变异程度不同，为了消除量纲影响和变量自身变异大小和数值大小的影响，便于不同单位或量级的指标能够进行比较和加权，数据的标准化是将数据按照指定区间大小进行缩放。其中最典型的的就是 0-1 标准化（归一化）。本文利用 0-1 标准化对原始数据线性变换，使结果落到[0, 1]区间，以降低各数据的数量级和差值。转换函数如下：

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (10)$$

其中 x_{max} 为样本数据的最大值， x_{min} 为样本数据的最小值。

4.5.3 人工神经网络模型的建立

人工神经网络具有自学习功能。把许多不同的数据和数据是否是正常状态下输入人工神经网络，网络就会通过自学习功能，慢慢学会识别类似的数据。自学习功能对于预测有特别重要的意义。而且其具有高速寻找优化解的能力。因此，人工神经网络模型可以用来预测电器正常状态下各数据的未来数值，然后再将实测值与预测值进行简单的比较就可以得出实时电器的状态是否正常。

用 matlab 可求出神经网络下的预测数据，本文选用神经网络中的 BP 模型：节点之间的传递函数为 S 型函数，即：

$$f(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

各层节点的输出按下式计算：

$$y = f\left[\sum_{i=0}^n \omega_j x_i - \theta\right] \quad (12)$$

各个节点输入是由输入信息的权重 ω 和阈值 θ 连接的。式中， y 是节点输出， x_i 是节点接受的信息， ω_j 是相关连接权重， θ 为阈值， n 是节点数。

(1) 神经网络的自学习算法

网络的学习过程具体如下。

- 1) 给网络赋一组随机初始权值，一般 $0 < \omega_j < 1$ ，并使其不相等。
- 2) 将输入数据标准化，生成输入信号 $(x_0, x_1, \dots, x_{n-1})$ ，确定期望输入信号 $(d_0, d_1, \dots, d_{k-1})$ ，使 d_k 在-1 和 1 之间且互不相等。
- 3) 逐层计算神经网络的实际输出值，

$$y_i = f\left[\sum_{i=0}^{n-1} k_{ij} x_i - \theta_j\right] \quad (13)$$

$$z_k = f\left[\sum_{j=0}^{n-1} k_{jk} x_j - \theta_k\right] \quad (14)$$

$$(j = 0, 1, \dots, m-1; k = 0, 1, \dots, l-1)$$

- 4) 从输出层开始反向调整权重，调整公式为

$$k_{jk} + ZW_{ky_i} \rightarrow k_{jk}, (j = 0, 1, \dots, m-1; k = 0, 1, \dots, l-1) \quad (15)$$

$$k_{ij} + ZW_{jx_i} \rightarrow k_{ij}, (j = 0, 1, \dots, m - 1; i = 0, 1, \dots, n - 1) \quad (16)$$

式中, Z 为学习因子, W 为误差修正量。为了使 Z 取值足够, 又不致产生振荡, Z 通常用调整公式

$$k_{ji}(t + 1) = k_{ji}(t) + ZW_{kj} + T[k_{ji}(t) - k_{ji}(t - 1)] \quad (17)$$

式中, T 为势态因子。计算期望值和计算值之间的误差, 若达到要求, 计算结束, 否则, 转到第 3 步继续计算, 直至达到要求。

(2) 神经网络在程序上的实现

依照人工神经网络的原理, 我们用 MATLAB 编制神经网络程序对数据进行了一些预测, 我们取工作环境 B 中正常的工作的参数 7 的前五组数据, 利用前五组数据对第六组数据进行预测, 运用的原理是输入前三组数据, 输出第四组, 再输入接下来的三组数据, 再预测第五组, 依此类推, 进行训练, 训练若干次后, 便可用三组数据来预测接下来未知的数据。我们用程序运行的第六组数据结果为 355.9162 (每次预测结果可能不同), 与实际测量值的 356 很接近, 证明数据满足正常变化趋势, 则电器运转正常。如若预测结果与实际测量值偏差过大, 可认为电器运转故障。

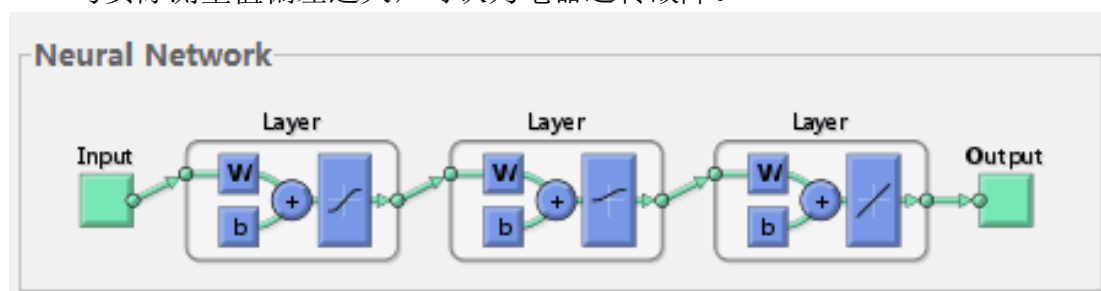


图 4-6 神经网络分析图

5 模型的评价与改进

5.1 模型的评价

5.1.1 模型的优点

1. 模型运用肖维勒准则以及小波分析方法对于大量异常数据进行了剔除处理, 减少了计算的复杂度。
2. 该模型运用了聚类分析的方法较好地解决了大量参数难以处理的问题, 并引入了故障判别半径这一参数, 对于分析与判别电器故障与否有较大的指导意义。
3. 运用人工神经网络模型对于数据进行预测, 通过预测值与实际测量值的比较来对电器故障进行判别, 使得模型更加新颖。

5.1.2 模型的缺点

1. 剔除异常数据时可能会对一些正常波动的数据进行误剔除, 影响模型对于电器故障的判断。
2. 因所建模型中的故障判别半径是依据经验进行更新的, 因此, 在没有大量实验数据的支撑下, 模型对于电器是否故障易出现误判的情况。

5.1.3 模型的改进

由于模型需要大量实验数据支持,所以可采用其他方法能够使模型对少量的数据处理后也能做出相对正确的判断。对于在数据处理过程中使用了均值进行计算,易出现被异常数据影响的情况,所以可以用其他指标来确定故障判别半径,排除异常数据的影响。

6 参考文献

- [1]许丽佳. 电子系统的故障预测与健康管理技术研究[D]. 电子科技大学, 2009.
- [2]刘旭. 数据挖掘在风力发电机组故障诊断中的应用[D]. 中国地质大学(北京), 2016.
- [3]姜启源, 谢金星, 叶俊. 数学模型. 第4版[M]. 高等教育出版社, 2011.。
- [4]牛文全, 李靖. 降水量的BP人工神经网络预测模型及其应用[J]. 西北农林科技大学学报:自然科学版, 2001, 29(4):000103-106.
- [5]杨春波. 基于灰色模型与人工神经网络的改进组合预测模型及其应用研究[D]. 山东师范大学, 2009.

附录

肖维勒准则对数据处理的 MATLAB 程序:

```
clc;

clear all;

%导入参数数据
a=xlsread('D:\1.xlsx');

[m,n]=size(a);

%对筛选出的 17 组参数进行处理
for j=1:17

    e=ones(m,1);

    omega=1+0.4*log(3563);

    average=mean(a(:,j)); %求平均值

    sigma=std(a(:,j)); %求标准差

    outliers(:,j)=abs(a(:,j)-e*average)>omega*e*sigma; %肖维勒判断

end

for i=1:m

    for j=1:n

        if outliers(i,j)==1

            a(i,j)=nan;

%异常的数据用 NAN 取代

        end

    end

end

[x,y]=find(isnan(a)==1);

a(x,:)=[];
```

小波分析去噪的 MATLAB 程序:

```
clc;
```

```

clear all;

%导入参数数据
a=xlsread('D:\1.xlsx');
data=a(:,2);
figure,plot(data);
xlabel('(s)');
ylabel('变量y');
title('原始数据');

%为角度赋初始值
a=pi/8;
b=pi/8;

%为重构滤波器h0(n)冲激响应赋值
h0=cos(a)*cos(b);
h1=sin(a)*cos(b);
h2=-sin(a)*sin(b);
h3=cos(a)*sin(b);
construct=[h0,h1,h2,h3];
frequency=4; %设定滤波器长度
decompose=construct(end:-1:1); %确定h0(-n)，分解滤波器

%导入参数数据
a=xlsread('D:\1.xlsx');
data=a(:,2);
figure,plot(data);
xlabel('时间(s)');
ylabel('变量y');
title('原始数据');
figure(1);
plot(data);
title('原数据');

%h0(n)性质校验

```

```

check1=sum(decompose);
check2=norm(decompose);
frequency=conv(data,decompose);%卷积
frequency_down=dyaddown(frequency);%信号低频细节
figure(2);
plot(frequency_down);
title('小波分析的低频系数');

```

运用人工神经网络对数据预测的 MATLAB 程序：

```

clc;

clear all;

%由工作环境B的正常数据的参数7的前五组预测第六组数据
%输入前五组数据及其相应训练结果

P =[356 356 356;356 356 360;356 360 356;360 356 356;356 356 356];

T = [360 356 356 356 356];

pmax = max(P);

pmax1 = max(pmax);

pmin = min(P);

pmin1 = min(pmin);

%归一化处理

for i=1:5

    P1(i,:)=(P(i,:)-pmin1)/(pmax1-pmin1);

    T1(i)=(T(i)-pmin1)/(pmax1-pmin1);

end

net = newff(P1',T1,[3 7],{'tansig' 'logsig'},'traingd');

%设置训练参数

for i = 1:5

    net.trainParam.epochs = 15000;

    net.trainParam.goal=0.01;

```

```
net.trainParam.lr = 0.1  
  
net = train(net,P1(i,:) ',T1(i));  
  
end  
  
y=sim(net,[356 356 356] ');  
  
y1=y*(pmax1-pmin1)+pmin1; %反归一化处理
```