

家用电器故障实时检测的建模算法研究

秦 强¹ 韩恩栋¹ 张慧雯²

(1. 西安交通大学 电子与信息工程学院 陕西 西安 710049;

2. 西安交通大学 人居环境与建筑工程学院 陕西 西安 710049)

[摘要]通过计算某种电器各参数在正常工作环境和故障环境下的距离以及分析各参数根据时间变化的图像,提出了一种基于距离和短周期、利用非线性综合加权法对电器运行状况进行实时监测并判别的方法,为用户提供家电运行实时状况,提升用户的使用体验。

[关键词]降维分析;故障相关性分析判别;短周期;非线性加权

[中图分类号]O232

[文献标识码]A

[文章编号]1671-5330(2017)05-0019-04

DOI:10.16140/j.cnki.1671-5330.2017.05.005

1 引言

家用电器是人们生活中不可或缺的一部分,其质量问题受到了广泛的关注。由于使用时间等多方面因素影响,电器会出现老化现象,之后其工作能力将呈现衰减趋势。当用户提出电器需要维修时,电器的故障情况通常来说已经非常严重,所以非常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保证用户的体验效果。本文根据某家电公司提供的经过脱敏处理后的某种电器运行数据,利用数学建模的方法来解决以上问题。

2 模型的建立与求解

2.1 模型的假设

1. 假设该家电 24 小时均为开机状态。

2. 假设在计算对应参数间的距离时,删除多余的行数对故障判别没有影响。

3. 假设不同的环境状态仅影响此家电部分参数的最值,即参数随时间变化的波形图中的波峰波谷值。

2.2 符号说明

x_{ij} : 矩阵 X 第 i 行第 j 列元素; S_{ij} : 矩阵 X 第 j 列元素的标准差; W_j : 第 j 个参数的权重; dis_j : 第 j

个参数的欧氏距离; y : 故障判别距离。

2.3 模型的建立与求解

为了处理传感器读取到的异常数据,首先进行数据检测,即根据人工观察及绘制数据变化曲线,找出曲线中明显的异常点,认为其为传感器读取到的异常数据。接下来利用差值等方法对发现的异常数据点进行修正。

对于某家电公司提供的经过脱敏处理后的某种电器运行数据,观察对比已知工作环境 A 正常及有故障的数据表,发现参数 8、9、10 中存在相似的较为显著的独立异常数据,例如在故障条件下,0:41:44 存在明显的突变异常数据,则对其进行删去处理。工作环境 A(有故障)中的参数 83 与其他各表中参数 83 数值明显不是同一数量级,判断其为错误数据,因此对其进行修正。

工作环境 A(正常)中的参数多达 96 个,且数字信息和文字信息相互掺杂,若直接对其进行分析会导致工作繁琐且难度较大,因此对已有数据进行恰当的处理具有重要意义^[1]。首先,筛选出一个小时内没有发生变化的参数,这些参数对相关性分析与故障判别无贡献,将其删去。接下来

[收稿日期]2017-08-06

[作者简介]秦强(1997-),男,山西太原人,主要从事计算机科学与技术研究;韩恩栋(1996-),男,河南安阳人,主要从事计算机科学与技术研究;张慧雯(1998-),女,新疆博乐市人,主要从事环境科学与技术研究。

将其余参数中的文字信息转化为数值,具体转化对照表如下:

名称	OFF	正常	故障	限	空
转换的数值	0	1	-1	0	0
名称	开	关	超高速	高速	中速
转换的数值	1	0	1	0	-0.5
名称	不屏蔽	设定	控制	普通	无
转换的数值	1	0	1	1	0

接下来使用 SPSS 软件对数据进行降维分析处理,得到旋转成分矩阵。筛选并标记出各组件内部大于 0.7 的值,作为此矩阵的主成分。对题目中不同工作环境与不同状态的数据均做此处理,分析出各环境中对应组件的相同主成分。

各组件的主成分表如下:

编号	主成分
组件 1	参数 1, 参数 3, 参数 4, 参数 5, 参数 10, 参数 18, 参数 60, 参数 72, 参数 73, 参数 80
组件 2	参数 8, 参数 83
组件 3	参数 74, 参数 75
组件 4	参数 6
组件 5	参数 2

使用 SPSS 软件对各组件的主成分进行相关性分析,得到 Pearson 相关系数矩阵,并且通过对矩阵的分析得到其显著性均小于 0.01,即相关性是显著的。

由于各参数所含信息量不同,故不同参数对于故障判别的重要程度也不尽相同,而标准差能够刻画一组数据的波动程度从而反映数据的信息量。所以根据矩阵各列向量的标准差给每个距离赋权重^[2]。

各参数的权重如下表:

参数 1	参数 2	参数 3	参数 4	参数 5
0.0511	0.0645	0.0436	0.0511	0.0407
参数 6	参数 7	参数 8	参数 9	参数 10
0.0620	0.0607	0.0632	0.0538	0.0299
参数 18	参数 60	参数 72	参数 73	参数 74
0.0406	0.0457	0.0514	0.0513	0.0612
参数 75	参数 78	参数 80	参数 83	
0.0612	0.0630	0.0449	0.0601	

使用 SPSS 22.0 分析参数的相关性得知,剩

下的 19 个参数 Pearson 相关性数值均不为 0,即不存在绝对独立的参数。因此,为了尽量避免信息的重复利用,文章使用非线性加权综合法建立故障判别模型。利用以下公式计算加权后的故障判别距离 y :

$$y = \prod_{j=1}^{19} dis_j w_j$$

计算后得到工作环境 A 的故障判别距离是 77.3999,利用相同方法对工作环境 B 下的数据进行处理,得到其故障判别距离为 72.8011,二者较为相近,说明所建立的模型通过了题目中的数据检验,认为其能够较好地反映真实情况,具有可操作性。

计算后得到工作环境 A 及工作环境 B 的故障判别距离均值为 75.1005。通过所建立的模型计算,得到工作环境 C 的故障判别距离为 75.7785,在数值上极为相近,说明其依然适用于题目所给数据。

上文所建立的模型,其特点是需要同一工作环境下的正常数据作为对照,否则将无法对数据进行故障判别距离的计算,从而无法判断电器的运行状态的正常与否。因为没有给出工作环境 D 与工作环境 E 的正常运行状况的数据,因此无法直接使用模型来判别电器状态为故障或正常。于是对模型进行以下的修正。

观察数据,发现前文没有考虑到电器的实时监测数据可能随时间发生变化。故此处加入对时间的考量。首先将不同工作环境下正常数据的相应参数随时间的变化图像分别绘制在同一坐标系下,观察图像的规律。此处以参数 6 的图像作为示例分析,作图 1。

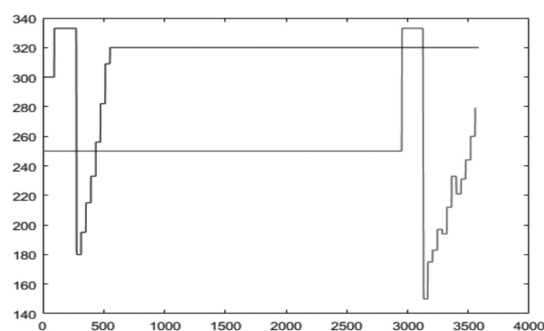


图 1 工作环境 A 与 B(正常)参数 6 的图像

观察发现剩余参数中,其中 6 个参数的图像在不同工作环境下均会出现形状大致相似的波峰与波谷,图像的差别主要在于波峰波谷的高度及其出现的时间。联系生活常识可知,不同的工作

环境可能会导致电器的物理参数相应地升高或降低,所以此处可以假设工作环境仅改变波峰高度及波谷深度。

由已知数据可知,实时监测的数据记录均从某一个整点时刻开始,到下一个整点时刻,经历一小时结束,但是所给出的时间段并不相同。结合考虑不同工作环境下均出现相似形状的波形图这一事实,可推知,在此电器常开的正常状态下,参数的数值变化呈现出一定的周期性,且周期较长。观察各参数的图像,推知这些周期均大于 1 小时。所以在不同工作环境下,虽然均从整点时刻开始检测数据,但出现波峰的时间却各不相同。因此可以假设此电器是常开的,实时监测表仅截取了该电器运行过程中的某一段时间的数据。

接下来观察各工作环境(有故障)中的各参数随时间变化趋势,下图以参数 10 为例。

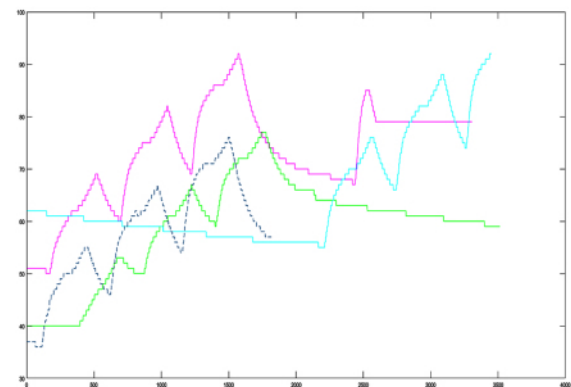


图 2 各工作环境下故障数据的参数 10 图像

经观察对比发现,故障表中某些参数相较于正常状态,同样呈现出周期性,但周期性明显缩短,因此如上图示,在 1 小时内可以观察到多个波峰相继出现。对 19 个参数依次分析,参数 3,参数 5,参数 6,参数 7,参数 10,参数 18 均出现了明显的短周期特征。因此,在同一工作环境中缺失正常状态的数据时,参数的短周期特征将作为判别故障与否的重要标志。

下面给出计算参数短周期的算法:

从前到后扫描参数的实时监测数据,若

$$A(i) > A(i-1) \text{ and } A(i) > A(i+1)$$

$A(i)$ 则为此参数的一个局部极大值,其中 $A(i)$ 为某个参数的第 i 个值,这样理论上可以找到一个小时内的各局部极大值,而局部极大值出现的时间差即为短周期。为了保证算法的严谨性,对相同的数值采用“缩点法”消除重复值。

从前到后扫描参数 A ,若 $A(i) \neq A(i+1)$ 则

有 $k = k + 1, A'(k) \leftarrow A(i)$,其中 A' 是进行缩点处理之后的向量。若当前元素与其后继数值不同,则缩点后的向量长度加 1,并将当前元素值赋给 $A'(k)$ 。记录下此值的初始位置: $\text{pos}(k) \leftarrow i$ 。进行缩点处理之后,部分图像仍存在一些噪点,若将其直接带入模型计算存在一定困难,需要通过人工干预取出合适的点计算时间差。

再次对 A' 搜索局部极大值,进而计算出在故障状态下参数 3,参数 5,参数 6,参数 7,参数 10,参数 18 的短周期的平均长度如下:

参数	参数 3	参数 5	参数 6	参数 7	参数 10	参数 18
短周期/s	532.33	532.33	533.25	533.50	527.00	533.25

将此表作为故障判别的参照表,通过考察监测数据中的以上六个参数是否具有短周期,判定此电器是否出现故障。

3 模型的检验与判断

下面检验模型的正确性。使用以上算法,求解出工作环境 C(故障)的相应参数的短周期,其值与参照表的相对误差如表所示:

参数	参数 3	参数 5	参数 6	参数 7	参数 10	参数 18
时间差/s	528	539	525	527	520	538
相对误差	0.81%	1.25%	1.55%	1.22%	1.33%	0.89%

可以看出,此时的时间差已经出现显著的短周期特征。由此可以得出结论,此模型对于检验工作环境 C(故障)是正确的。

作图发现工作环境 C(正常)的以上参数没有出现短周期,因此该模型也适用于判断工作环境 C(正常)的情况。

类似地,求解工作环境 D 中相应参数的时间差:

参数	参数 3	参数 5	参数 6	参数 7	参数 10	参数 18
时间差/s	532	532	531.5	561.5	524.5	532
相对误差	0.06%	0.06%	0.33%	5.25%	0.47%	0.14%

可以明显观察到,工作环境 D 的时间差出现短周期的特征,据此可判定工作环境 D 的数据是有故障的。

经计算,工作环境 E 的以上参数没有出现短周期特征的时间差,绘制个别参数图像如下:

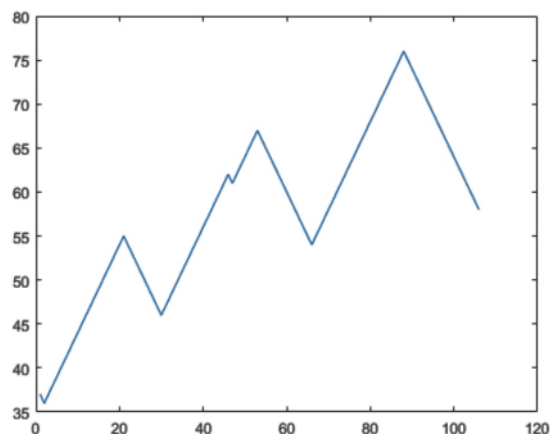


图3 工作环境 E 的参数 3

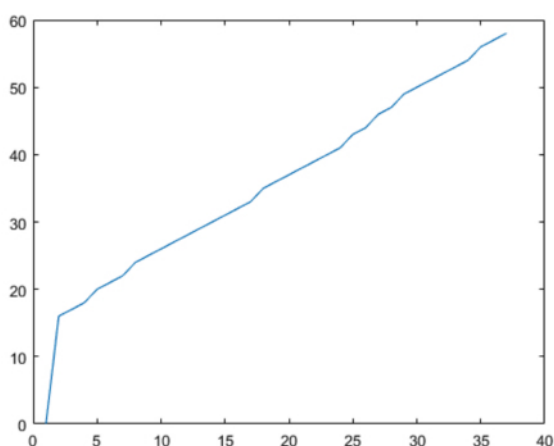


图4 工作环境 E 的参数 10

因此判定工作环境 E 的数据是正常的。

根据实际情况,将故障判断为正常会影响用户的正常使用,而将正常判断为故障对用户的正常使用影响较小。因此将故障判断为正常的误判就更应该避免。

当一组数据具有短周期的特征,但周期却比参照表中短周期时间长出许多时,即认为进入了判定的“灰色地带”。此时极易出现误判。

由于故障状态具有将参数的周期向缩短的方向引领的趋势,所以当周期较短时,模型可以容易地判断为故障,即不易发生误判^[3]。

但是对于进入“灰色地带”的周期,其数值虽显著大于参照表中对值,但相较于正常周期长度仍有一定差距,更容易产生误判,因此嵌入影响因子以明确较长周期的归属。

若此六个参数中的某些参数具有被模型判断为故障的趋势,则适当增大这几个参数的权重。影响因子为此六个参数的周期的动态权重。此处动态权重使用分段变幂函数来确定。即将参数的周期分为 K 级,不同的级别由下式计算不同的权重:

$$w(i, x) = x^{\frac{1}{k}}$$

$$k = 1, 2, \dots, K$$

$$i = 1, 2, \dots, 6$$

$w(i, x)$ 为第 i 个参数的周期 x 的动态权重, k 为此参数的周期的级别。

这样,再利用非线性加权综合法求出综合周期:

$$T = \prod_{i=1}^6 t(i) w(i, x)$$

$t(i)$ 是第 i 个参数的周期。

级别区间的确定可根据实际情况从参照表的数值到 1 小时作非线性划分。

4 结束语

文章未全部考虑检测得到的 96 个参数,通过分析筛选只选择了其中 6 个故障判别参数进行分析,从而简化了工作量。其次,故障判别模型分析了 6 个故障判别参数随时间变化的图像的周期,计算过程简单,图像直观,操作较为简便。但是若某些环境下此电器出现故障,且故障类型与所给数据均不相同,即引发除文章所讨论的 6 个故障判别参数以外的参数的变化,则本模型可能无法做出精确判断。其次本模型对数据中缺失时间段的数据处理不到位,可能对故障判别存在影响。

[参考文献]

- [1] 周俊临. 基于数据挖掘的分布式异常检测[D]. 电子科技大学 2010.
- [2] 邓锦军. 基于欧拉距离法的制冷压缩机故障判断[J]. 家电科技 2008(19).
- [3] 孟静. 异常数据挖掘算法研究与应用[D]. 无锡: 江南大学 2013.

[责任编辑: 张怀涛]