

DOI: 10.13382/j.jemi.B2003062

小样本数据处理的加速寿命预测方法*

佐磊^{1,2} 胡小敏¹ 何怡刚^{1,2,3} 孙洪凯¹ 李兵^{1,2}

(1. 合肥工业大学 电气与自动化工程学院 合肥 230009; 2. 合肥工业大学 可再生能源接入电网技术
国家地方联合工程实验室 合肥 230009; 3. 武汉大学 电气与自动化学院 武汉 430072)

摘要: Bayes Bootstrap 法在小样本预测领域应用成熟广泛,但由于其随机产生的自助样本中存在不利于预测精度的野值点造成预测偏差较大,针对此不足,提出 Bayes Bootstrap & k-means 方法。在拥有小样本失效数据情况下,首先采用 Bayes Bootstrap 法产生自助样本对原有寿命数据进行容量扩充,再采用 k-means 方法对其进行数据聚类分析,尽可能去除野值点,筛选出更加符合预测规律的数据点。最后以多芯片组件互连结构双应力加速寿命进行预测为例计算验证了该方法相比仅采用 Bayes Bootstrap 法,预测精度提高了约 81.44%,有一定的工程意义。

关键词: 小样本失效数据; 数据处理; Bayes Bootstrap & k-means; 寿命预测

中图分类号: TN602; N941.5 **文献标识码:** A **国家标准学科分类代码:** 510.30

Accelerated life prediction method for small sample data

Zuo Lei^{1,2} Hu Xiaomin¹ He Yigang^{1,2,3} Sun Hongkai¹ Li Bing^{1,2}

(1. School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China;
2. National Local Joint Engineering Laboratory of Renewable Energy Access to Power Grid Technology,
Hefei University of Technology, Hefei 230009, China;
3. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China)

Abstract: The Bayes Bootstrap method is widely used in the field of small sample prediction. However, due to the random value points that are not conducive to the prediction accuracy in the randomly generated self-service sample, the prediction deviation is large. In view of this deficiency, this paper proposes the Bayes Bootstrap & k-means method. In the case of having small sample failure data, use the Bayes Bootstrap method to generate self-service samples to expand the capacity of the original life data firstly, and then use the k-means method to perform data clustering analysis to remove outliers as much as possible and filter out more data points that meet the forecasting rules. Finally, the multi-chip module interconnection structure double stress accelerated life prediction is used as an example to verify the calculation method. Compared with the Bayes Bootstrap method, the prediction accuracy is improved by about 81.44%, which has certain engineering significance.

Keywords: small sample failure data; data processing; Bayes Bootstrap & k-means; lifetime prediction

0 引言

目前,针对电子设备寿命预测的方法已有诸多文献进行研究^[1-3],其中多芯片组件(multi-chip module, MCM)

作为一种先进的电子产品具有高可靠度、长寿命特征,其寿命试验数据一般具有小样本性。已有的神经网络法针对大样本时预测精度高,但小样本情况下预测精度存在不足^[4];灰色预测方法在小样本条件下做寿命预测运用广泛,但其在对非时序数据的预测应用受到了一定

收稿日期: 2020-04-09 Received Date: 2020-04-09

* 基金项目: 装备预先研究重点项目(41402040301)、国家重点研发计划(2016YFF0102200)、国家自然科学基金重点项目(51637004)、国家自然科学基金(51777050, 51577046)资助项目

限制^[5]。

Bayes Bootstrap 被广泛应用在小样本数据的预测领域,此方法针对小样本,在给定的原始样本的基础上采用计算机抽样产生新的样本来增大样本容量,从而更好地进行参数估计^[6],但其预测精度会受到在随机抽样过程中产生的野值点影响。近年来,大量学者研究如何对其进行改进或结合其他方法对扩充后的数据进行处理,从而提高预测精度^[7-8]。周凯等^[9]结合形态相似距离对 Bayes Bootstrap 法扩容后的数据进行曲线拟合做预测估计,该方法在数据扩容基础上有效抑制了“野值点”降低拟合精度的弊端,但需要同时计算各组数据的多种距离值,过程较为复杂。汪晓洋^[10]利用 aiNet 对 Bayes Bootstrap 扩充后的样本聚类分析,也有效解决了 Bayes Bootstrap 方法的预测精度不足问题,但免疫算法中“疫苗”的获取过程较为繁琐,适用范围具有一定的局限性。为进一步探索结合其他方法弥补 Bayes Bootstrap 的不足,本文在详细叙述了 Bayes Bootstrap 数据处理过程的基础上结合适用范围广的 k-means 法提出了 Bayes Bootstrap&k-means 寿命预测方法,其中 k-mean 算法具有聚类效率高、效果好等优点^[11-12],能够快速实现剔除野值点,减小预测误差,所提方法对具有小样本性、失效机理与结构复杂的 MCM 应用效果好,且易实现。结合案例进行阐述,利用 MATLAB 软件编程,验证了结合 k-means 后预测精度确有提高,证明了此法是有效且适用的,有一定的工程意义。

1 热-电双应力加速模型

MCM 中的互连结构是影响其整机寿命的最重要环节^[13],除考虑影响寿命的结构因素外,复杂的服役环境需要通过试验对失效的不确定因素进行有效的把握和预防,因此,基于试验条件下 MCM 的寿命研究产品的工程应用中具有重要意义。

导致 MCM 失效的主要应力为环境应力,针对电子器件的寿命预测要考虑的因素众多,仅凭借单一因素对电子器件做出的寿命预测并不是很准确^[14]。本文考虑在恒定加速寿命试验方案下同时选取热应力和电应力作为加速应力。用 Peck 加速模型来描述寿命与热应力和非热应力(常为电应力)的关系,能够有效弥补仅考虑单一应力作用时的不足,其特征寿命在热-电恒加试验时满足关系式:

$$t_e = AS^{-n} \exp\left(\frac{E_a}{kT}\right) \quad (1)$$

式中: A 是常数,其值与材料相关, $k=0.8617 \times 10^{-4} \text{ eV/K}$; T 和 S 分别为开氏温度和电应力; E_a 是激活能, n 是电应力加速因子, E_a 和 n 是计算产品寿命的重要待定参数,其值与材料和失效机理密切相关。

根据式(1)得出热应力加速因子和电应力加速因子:

$$AFT = \exp\left[\frac{E_a}{k} \left(\frac{1}{T_{op}} - \frac{1}{T_{str}}\right)\right] \quad (2)$$

$$AFS = \left(\frac{S_{str}}{S_{op}}\right)^n \quad (3)$$

得出其他应力环境下的特征寿命为:

$$t_{eop} = t_{estr} \cdot AFT \cdot AFS \quad (4)$$

对式(1)两边同时取对数得:

$$\ln t_e = -n \ln S + B \quad (5)$$

$$\ln t_e = \frac{E_a}{kT} + C \quad (6)$$

其中:

$$B = \ln A + \frac{E_a}{kT} \quad (7)$$

$$C = \ln A - n \ln S \quad (8)$$

可见,此加速模型下当双应力中某一应力不变时,特征寿命的对数总与变化应力的已知函数总呈线性关系。两式可化简为:

$$\ln t_e = a + b\varphi(S) \quad (9)$$

式中: S 为双应力中任一加速应力; $\varphi(S)$ 为应力 S 的已知函数; a, b 为常数。

利用式(9)研究 MCM 在一定失效机理情况时的各加速应力水平下的特征寿命之间的换算关系。本文就如何提高经线性拟合对参数 E_a 和 n 进行估计的精度而提出了新的 Bayes Bootstrap&k-means 方法。

2 Bayes Bootstrap&k-means 方法

2.1 Bayes Bootstrap 的基本原理介绍

Bayes Bootstrap 又称自展法,是用小样本估计总体值的一种非参数方法,此方法的思想,对初始数据进行有放回的抽样,进而生成一系列伪样本,通过对伪样本的计算,获得统计量的分布。文献[10]指出,在原始样本数量足够多时 Bayes Bootstrap 方法能够无偏地接近总体分布。

1) 基本原理

在工程研究中,假设某种产品的性能参数 X 独立同分布,样本数据 $X_i \sim N(\mu, \sigma^2)$, $i=1, 2, \dots, n$, n 为试验的样本总数^[15]。性能参数评估的关键是 μ 和 σ 的确定,所以如何提高其估计精度对寿命预估具有重要意义。

对于来自总体的原始样本 $X = (x_1, x_2, \dots, x_n)$, 样本的分布函数 $F(X)$ 未知,样本 X 中的某一统计量为 $\theta(F)$, $\hat{\theta}(F)$ 为 $\theta(F)$ 的估计量,则其估计偏差 T_n 可表示为:

$$T_n = \hat{\theta}(F) - \theta(F) \quad (10)$$

若从 $F(X)$ 中从新抽样得到新的自助样本服从分布

函数 $F^*(X)$, 则有:

$$R_n^* = \hat{\theta}(F) - \hat{\theta}(F^*) \quad (11)$$

则可以用 R_n^* 的分布情况逼近 T_n , 多次重复抽样得到模拟样本, 且对于每次抽样所得的 R_n^* , 利用式(11)计算出统计量的每个可能值, 将模拟样本作为 $\theta(F)$ 的样本, 采用统计理论求 $\theta(F)$ 的相关值。

2) 具体步骤

1) 构造随机变量序列 $V_i = U_i - U_{i-1} (i = 1, 2, \dots, n)$, 其中 $n - 1$ 个随机数 U_1, U_2, \dots, U_{n-1} , 均匀分布且 $0 \leq U_i \leq 1, U_0 = 0, U_n = 1$ 。

(1) 样本均值为:

$$\theta_\mu = \sum_{i=1}^n V_i X_i \quad (12)$$

3) 样本方差为:

$$\theta_s = \sum_{i=1}^n V_i (X_i - \sum_{j=1}^n V_j X_j)^2 \quad (13)$$

4) 重复步骤 1)~3), 直到式(12)、(13)计算结果的均值达到稳定为止, 从而得出均值和方差的估计。

2.2 k-means 数据聚类方法简介

1) k-means 数据聚类方法的起源与发展

k-means 算法是聚类分析方法中常被使用的一种迭代求解的无监督学习算法, 也被称为 k-means 算法, 此聚类算法主要采用距离作为相似性的评价指标, 两个数据点的间距越小时表明两点相似度越大。k-means 算法是通过多次重复迭代把多个对象分成 k 个不同的类簇, 使之具有较高的簇内相似度, 较低的簇间相似度。由于 k-means 算法聚类效率高、步骤简单快速、聚类效果好且能够处理无标签数据, 被广泛应用于大规模低维数据集上的聚类^[16-17]。

k-means 算法的处理过程如下: 首先, 将待处理数据任意分为 k 个类, 类用于表示初始类簇的中心(即类簇平均值); 计算数据与各簇中心的距离, 然后将其赋给各自距离最近的类簇; 再重新计算每个簇的平均值。这个过程不断重复, 直到准则函数收敛。一般情况下采用误差平方和准则, 其函数公式为:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (14)$$

式中: E 为待处理数据的平方误差总和; p 是空间中的点; m_i 是簇 C_i 的均值。该目标函数最常使用欧氏距离用于度量距离, 使生成的簇尽可能紧凑独立。

k-means 算法对高维数据存在计算速度缓慢等不足, 因此有大量文献对其进行优化或者采用其他算法替换, 但本文中的数据维度低, 因此适用。

2) k-means 算法的具体步骤

随着人们对 k-means 算法研究的逐渐深入, 其应用范围十分广泛, 涉及到优化、控制、通信等多学科领域。

k-means 算法在解决数据聚类分析、数据挖掘以及数据浓缩和任务归类等方面应用成熟, 已受到学术界的广泛认可。k-means 算法的一般步骤^[18-19]如图 1 所示。

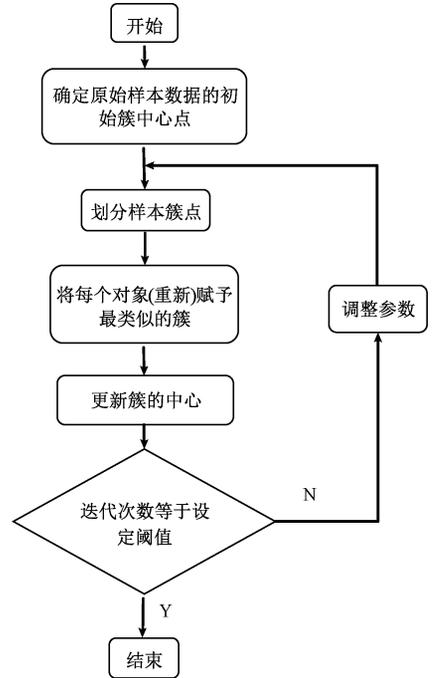


图 1 k-means 算法的一般步骤

Fig. 1 k-means algorithm general steps

(1) 任意选择 k 个类别对原始样本数据处理, 确定初始的簇中心;

(2) 划分样本簇点, 利用式(14), 根据每个故障数据的聚类样本的中心点, 计算出每个故障信号样本数据与这些中心样本参数之间的距离, 并且根据最小距离将相应的故障信息数据重新进行划分, 将各个样本数据簇的点划分成距离其比较近的中心所表示的样本簇中;

(3) 根据各簇的平均值, 将每个样本数据(重新)聚类到最相似的簇;

(4) 重新计算各簇的平均值, 更新簇的中心;

(5) 判断迭代次数是不是等于设定阈值, 若迭代次数不等于设定阈值, 则调整参数, 回到步骤(2), 重复进行步骤(2)和(3), 继续进行迭代计算, 若达到设定阈值, 迭代计算结束。

3 实验结果及分析

以多芯片组件某引线互连的热-电双应力加速 Peck 模型为例, 假设其中失效参数如下: $\ln A = 14, n = 1.3, E_a = 0.6$, 当加速温度统一在 $T = 473 \text{ K}$, 电应力(此处设为电流密度)不同, 分别取 $S_1 = 4 \times 10^5 \text{ A/cm}^2, S_2 = 6 \times 10^5 \text{ A/cm}^2, S_3 = 8 \times 10^5 \text{ A/cm}^2$ 时, 进行实验。

1) 真实试验数据模拟

由式(7)得出:

$$B = \ln A + \frac{E_a}{kT} = 14 + \frac{0.6}{0.8617 \times 10^{-4} \times 473} = 28.72$$

$$\ln t_{es_1} = -1.3 \ln 400\,000 + 28.72 = 13.2409,$$

则:

$$t_{es_1} = e^{13.2409} = 562\,924\text{ s} = 156.3678\text{ h}$$

$$\ln t_{es_2} = -1.3 \ln 600\,000 + 28.72 = 12.7544$$

则:

$$t_{es_2} = e^{12.7544} = 346\,071\text{ s} = 96.1309\text{ h}$$

$$\ln t_{es_3} = -1.3 \ln 800\,000 + 28.72 = 12.4092$$

则:

$$t_{es_3} = e^{12.4092} = 245\,045\text{ s} = 68.0682\text{ h}$$

当 $S = 1 \times 10^5 \text{ A/cm}^2, T = 323 \text{ K}$ 时有:

$$AFT = \exp\left[\frac{E_a}{k} \left(\frac{1}{T_{op}} - \frac{1}{T_{str}}\right)\right] =$$

$$\exp\left[\frac{0.6}{0.8617 \times 10^{-4}} \left(\frac{1}{423} - \frac{1}{473}\right)\right] = 5.6977$$

$$AFS = \left(\frac{S_{str}}{S_{op}}\right)^n = \left(\frac{0.4}{0.1}\right)^{1.2} = 5.2780$$

得出此应力环境下的特征寿命为:

$$t_{eop} = t_{estr} \cdot AFT \cdot AFS =$$

$$156.3678\text{ h} \times 5.6977 \times 5.2780 \approx 0.5368\text{ years}$$

在对加速寿命模型的研究中,常用对数正态分布表示,由文献[20]可知因封装材料引起的失效符合对数正态分布,因而在恒定应力加速寿命试验时,应符合相应寿命分布的基本假设,因此,产品在 $S_1 = 4 \times 10^5 \text{ A/cm}^2, T = 473 \text{ K}$ 时,符合期望为 156.3678 h 的对数正态分布,即:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi} \times 156.3678} e^{-\frac{(x - \ln 156.3678)^2}{2\sigma^2}}$$

假设寿命的标准差 $\sigma^2 = 0.2$, 则可以在对数正态分布下随机抽取几个数作为寿命值来模拟产品在此应力下的试验寿命结果数据,如图 2 所示。同理,产品在电应力为 $S_2 = 6 \times 10^5 \text{ A/cm}^2$ 和 $S_3 = 8 \times 10^5 \text{ A/cm}^2$ 的对数正态分布分别以 96.1309 h, 68.0682 h 为期望,以 0.2 为标准差,生成对数正态分布并随机抽取 5 个数据,模拟数据结果汇总如表 1 所示。

表 1 对数正态分布模拟试验数据

Table 1 Lognormal distribution simulation test data (h)

$S_1 = 4 \times 10^5 \text{ A/cm}^2$	$S_2 = 6 \times 10^5 \text{ A/cm}^2$	$S_3 = 8 \times 10^5 \text{ A/cm}^2$
183.937 5	96.106 7	87.183 5
151.963 2	102.868 7	88.207 3
153.274 2	118.201 7	64.383 7
133.242 9	72.583 3	71.088 5
172.578 1	78.224 7	77.219 0

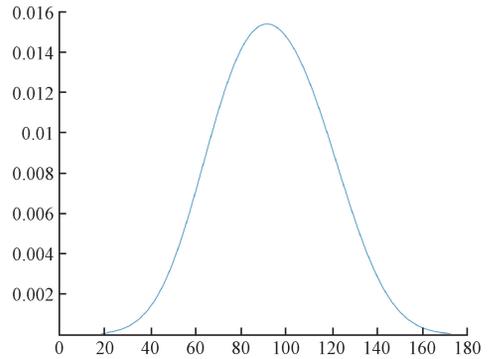


图 2 对数正态分布

Fig. 2 Lognormal distribution

2) 不经 Bootstrap&k-means 方法处理的寿命预测

由式(5)~(8)得出特征寿命的对数 $\ln t_e$ 与电应力的对数 $\ln S$ 、温度的倒数 $1/T$ 分别呈线性关系。

将表 1 的数据转化成秒后取对数,通过 MATLAB 软件对式(5)、(6)的 n 和 E_a 进行最小二乘法线性拟合估计,得到的结果为:

$$\begin{pmatrix} n \\ \frac{E_a}{kT} \end{pmatrix} = \begin{pmatrix} 1.048 \\ 12.7506 \end{pmatrix}, E_a = 0.52$$

当 $S_1 = 4 \times 10^5 \text{ A/cm}^2, T = 473 \text{ K}$ 时,有:

$$\ln t_{es_1} = -1.048 \ln 400\,000 + 14 + 12.7506 = 13.2322$$

$$AFT = \exp\left[\frac{E_a}{k} \left(\frac{1}{T_{op}} - \frac{1}{T_{str}}\right)\right] =$$

$$\exp\left[\frac{0.52}{0.8617 \times 10^{-4}} \left(\frac{1}{423} - \frac{1}{473}\right)\right] = 4.518$$

$$AFS = \left(\frac{S_{str}}{S_{op}}\right)^n = \left(\frac{0.4}{0.1}\right)^{1.048} = 4.275$$

得出 $S = 1 \times 10^5 \text{ A/cm}^2, T = 423 \text{ K}$ 应力环境下的特征寿命为:

$$t_{es} = t_{es1} \cdot AFT \cdot AFS =$$

$$155.0133\text{ h} \times 4.518 \times 4.275 \approx 0.342\text{ years}$$

3) 经 Bootstrap 方法处理的寿命预测

由于加速寿命分布服从对数正态分布,因此取对数后的数据均符合正态分布,对表 1 各应力条件下数据的期望和方差分别进行点估计。

对 3 种电应力分别建立期望 $\hat{\mu}_i$ 和方差 $\hat{\sigma}_i, i = 1, 2, 3$ 的对数正态分布函数式,并用 Bayes Bootstrap 方法分别在 3 条正态曲线上随机生成 500 个数为例扩充小样本容量,求得当仅用 Bayes Bootstrap 方法均值分别为 154.6702、101.9611 和 69.3763,求得参数 E_a 和 n 的估计值分别为:

$$\begin{pmatrix} n \\ \frac{E_a}{kT} \end{pmatrix} = \begin{pmatrix} 1.1483 \\ 14.0567 \end{pmatrix}, E_a = 0.5729$$

当 $S_1 = 4 \times 10^5 \text{ A/cm}^2, T = 473 \text{ K}$ 时有:

$$\ln t_{eS_1} = -1.1483 \ln 400\,000 + 14 + 14.0567 = 13.2445$$

$$AFT = \exp\left[\frac{E_a}{k} \left(\frac{1}{T_{op}} - \frac{1}{T_{str}}\right)\right] =$$

$$\exp\left[\frac{0.5729}{0.8617 \times 10^{-4}} \left(\frac{1}{423} - \frac{1}{473}\right)\right] = 5.2670$$

$$AFS = \left(\frac{S_{str}}{S_{op}}\right)^n = \left(\frac{0.4}{0.1}\right)^{1.1483} = 4.9130$$

得出 $S = 1 \times 10^5 \text{ A/cm}^2, T = 423 \text{ K}$ 应力环境下的特征寿命:

$$t_{eS} = t_{estr} \cdot AFT \cdot AFS =$$

$$154.6702 \times 5.2670 \times 4.9130 \approx 0.4569 \text{ years}$$

4) 经 Bootstrap & k-means 方法处理的寿命预测

利用 Bootstrap 方法随机生成的自助样本并不是都可以使用,一些跳脱的点不利于提高预测精度,所以有必要对扩充后的数据进行处理,本文结合 k-means 方法对扩充后的数据进行聚类分析,筛选出更加有利于提高参数估计的精度样本数据,再将原始试验得到的 5 个样本和 Bootstrap 方法得到的 500 个自助样本结合到一起,形成一个含有 505 个样本的向量,将此 505 个样本代入到 k-means 算法里进行数据聚类分析,数据的聚类分析结果如图 3 所示(以期望为 156.3678 h 为例),将目标分为 5 组时,得到各组数据结果如表 2 所示。

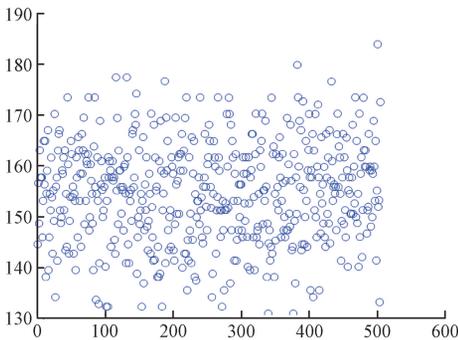


图 3 聚类效果

Fig. 3 Clustering effect graph

表 2 聚类结果 ($i = 1, S_1 = 4 \times 10^5 \text{ A/cm}^2$ 时)

Table 2 Clustering results ($i = 1, S_1 = 4 \times 10^5 \text{ A/cm}^2$)

第 1 组	聚类到样本数量共 129 个,均值为 148.2478,其中原始试验数据为 1 个
第 2 组	聚类到样本数量共 110 个,均值为 162.9569,其中原始试验数据为 0 个
第 3 组	聚类到样本数量共 71 个,均值为 138.9698,其中原始试验数据为 1 个
第 4 组	聚类到样本数量 49 个,均值为 171.7127,其中原始试验数据为 2 个
第 5 组	聚类到样本数量共 146 个,均值为 156.0168,其中原始试验数据为 1 个

由于原始试验得到的数据一般最能反映产品的寿命特征,因此优先保留含有原始试验数据的组,另外 k-means 聚类分析方法能够将不利于预测精度的数据分离出来,减小估计偏差。

由表 2 的聚类结果可知,第 2 组虽然均值靠近原始数据均值,但不含原始点数据,认为此组子样本的有效性较差,不可取;而第 1 组和第 3 组虽然含有 1 组原始数据,但其均值对比原始数据相差较大,也不可取,因此,取第 4 组和第 5 组的数据作为预测样本数据是最合理有效,最终从 505 组数据聚类筛选出 195 组,求出这 195 组数据的均值。同理对应力 $S_2 = 6 \times 10^5 \text{ A/cm}^2$ 和 $S_3 = 8 \times 10^5 \text{ A/cm}^2$ 分别筛选出 156 组和 426 组,并求得均值为 95.7939 和 70.35476。最终,利用最小二乘法求得 E_a 和 n 的估计值分别为:

$$\begin{pmatrix} n \\ E_a \\ kT \end{pmatrix} = \begin{pmatrix} 1.1904 \\ 14.6104 \end{pmatrix}, E_a = 0.5955$$

当 $S_1 = 4 \times 10^5 \text{ A/cm}^2, T = 473 \text{ K}$ 时有:

$$\ln t_{eS_1} = -1.1904 \ln 400\,000 + 14 + 14.6104 = 13.2552$$

$$AFT = \exp\left[\frac{E_a}{k} \left(\frac{1}{T_{op}} - \frac{1}{T_{str}}\right)\right] =$$

$$\exp\left[\frac{0.5955}{0.8617 \times 10^{-4}} \left(\frac{1}{423} - \frac{1}{473}\right)\right] = 5.6237$$

$$AFS = \left(\frac{S_{str}}{S_{op}}\right)^n = \left(\frac{0.4}{0.1}\right)^{1.1904} = 5.2083$$

得出 $S = 1 \times 10^5 \text{ A/cm}^2, T = 423 \text{ K}$ 应力环境下的特征寿命:

$$t_{eS} = t_{estr} \cdot AFT \cdot AFS =$$

$$158.6199 \times 5.6237 \times 5.2083 \approx 0.5304 \text{ years}$$

5) 结果分析

表 3 结果对比表

Table 3 Results comparison table

参数	真实结果	未 Bayes		Bayes		Bayes Bootstrap & k-means 处理	误差
		Bootstrap 处理	误差	Bootstrap 处理	误差		
n	1.2	1.05	0.13	1.15	0.04	1.19	8×10^{-3}
E_a	0.6	0.52	0.13	0.57	0.05	0.5955	7.5×10^{-3}
t_e	0.54	0.34	0.36	0.46	0.15	0.53	0.0119

对比表 3 的结果可知,针对小样本失效数据,直接用最小二乘法而不对小样本数据做任何处理时的参数估计偏差最大,仅经 Bayes Bootstrap 处理而不经 k-means 聚类筛选的数据偏差相比未经处理的小得多,但误差仍然较大,而经 Bayes Bootstrap & k-means 方法处理后数据偏差对比前两种方法明显减小,预测精度大幅度提高。

4 结 论

本文针对拥有小样本失效数据的寿命预测,提出 Bayes Bootstrap & k-means 方法。先采用 Bayes Bootstrap 法产生自助样本对原有寿命数据进行容量扩充,再采用 k-means 方法对其进行数据聚类分析,尽可能去除野值点,筛选出更加符合预测规律的数据点。以多芯片组件薄弱环节失效为例研究其在小样本失效数据条件下的寿命预测方法,计算验证了该方法的有效性。

由实例结果得出,现有 Bayes Bootstrap 方法虽然具有一定的适用性,但由于野值点的存在产生较大误差的缺点,针对这一不足,经 k-means 对扩容数据聚类分析后,有效筛选出更有利于减小预测误差的数据,提高了其预测精度。本例中相比未经 k-means 聚类处理的 Bayes Bootstrap 方法,总体精度大幅提高约 81.44%,预测精度又进一步提高。因此 Bayes Bootstrap & k-means 方法在对基于小样本数据的寿命预测的实际应用中可以进行推广。

本文的主要贡献如下:

1) 基于 Bayes Bootstrap 结合数据聚类分析方法提高预测精度之外提供了一种针对小样本失效数据条件下的寿命预测方法,此方法为失效数据少、寿命长、可靠性高以及结构复杂的电子产品的寿命预测提供了新思路。

2) 通过 k-means 方法对大量自助样本进行聚类分析,筛选出有利于提高预测精度的样本点,剔除野值点,进一步提高了预测精度,扩展了 Bayes Bootstrap 方法的适用范围。

参考文献

- [1] 姜媛媛,刘柱. 锂电池剩余寿命的 ELM 间接预测方法[J]. 电子测量与仪器学报, 2016,30(2): 179-185.
JIANG Y Y, LIU ZH. ELM indirect prediction method for the remaining life of lithium-ion battery[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(2): 179-185.
- [2] 张吉宣,贾建芳. 电动汽车供电系统锂电池剩余寿命预测[J]. 电子测量与仪器学报, 2018,32(3): 60-66.
ZHANG J X, JIA J F. Remaining useful life prediction of lithium-ion battery in electric vehicles power supply system [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(3): 60-66.
- [3] SU S N, AKKARA F J, THAPER R. A state-of-the-art review of fatigue life prediction models for solder joint[J]. Journal of Electronic Packaging, 2019, 141(4): 040802.
- [4] 金亮,王飞,杨庆新,等. 永磁同步电机性能分析的典型深度学习模型与训练方法[J]. 电工技术学报,

2018,33(1):41-48.

JIN L, WANG F, YANG Q X, et al. Typical deep learning model and training method for performance analysis of permanent magnet synchronous motor [J]. Transactions of China Electrotechnical Society, 2018, 33(1):41-48.

- [5] 吴汶捷,何怡刚. 油中溶解气体的灰色模型残差修正融合算法[J]. 电子测量与仪器学报, 2018, 32(10): 87-94.
WU W J, HE Y G. Residual-modified fusion algorithm of dissolved gas in transformer oil based on grey model[J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(10): 87-94.
- [6] 王晓玲, 谢怀宇, 王佳俊. 基于 Bootstrap 和 ICS_MKELM 算法的大坝变形预测[J]. 水力发电学报, 2020, 39(3): 106-120.
WANG X L, XIE H Y, WANG J J. Prediction of dam deformation based on Bootstrap and ICS-MKELM algorithms [J]. Journal of Hydroelectric Engineering, 2020, 39(3): 106-120.
- [7] 董金龙,汪立新,盛立昊. 基于改进 Bayesian_Bootstrap 陀螺仪寿命预测研究[J]. 压电与声光, 2014, 36(3): 384-388.
DONG J L, WANG L X, SHENG L H. Study on gyroscope life prediction based on improved bayesian bootstrap [J]. Piezoelectrics & Acousto-optics, 2014, 36(3): 384-388.
- [8] 张龙,叶松,王晓蕾. 基于灰自助模型的气压动态测量结果评估方法[J]. 仪器仪表学报, 2017, 38(7): 1645-1652.
ZHANG L, YE S, WANG X L. Evaluation method for dynamic measurement result of atmospheric pressure based on grey bootstrap model [J]. Chinese Journal of Scientific Instrument, 2017, 38(7): 1645-1652.
- [9] 周凯,丁坚勇,田世明,等. 基于小样本性能数据的电气设备可靠性评估与预测方法研究[J]. 电网技术, 2018, 42(6): 1967-1974.
ZHOU K, DING J Y, TIAN SH M, et al. Research on assessment and prediction of electrical equipment reliability based on small sample performance data [J]. Power System Technology, 2018, 42(6): 1967-1974.
- [10] 汪晓洋. 小样本下可靠性试验方法与数据处理的研究[D]. 成都:电子科技大学, 2012.
WANG X X. Research on reliability test methods and data processing in small samples [D]. Chengdu: University of Electronic Science and Technology of China, 2012.
- [11] 李金涛,艾萍,岳兆新. 基于 K-means 聚类算法的改进[J].

- 国外电子测量技术,2017,36(6):9-13,21.
- LI J T, AI P, YUE ZH X. Improvement of clustering algorithm based on K-means [J]. Foreign Electronic Measurement Technology,2017,36(6):9-13,21.
- [12] 邓滨玥. K 均值优化算法综述[J]. 软件, 2020, 41(2):188-192.
- DENG B Y. A survey on advanced K-means algorithm[J]. Computer Engineering & Software, 2020, 41 (2): 188-192.
- [13] 叶培莲. 基于失效物理的多芯片组件可靠性分析[D]. 成都:电子科技大学,2017.
- YE P L. Reliability analysis of multi-chip module based on physics of failure [D]. Chengdu: University of Electronic Science and Technology of China,2017.
- [14] 王春晖,李芳蓉,李晓钢. DC/DC 变换器加速寿命试验技术研究[J]. 环境技术, 2019(6): 136-140.
- WANG CH H, LI F R, LI X G. Research on accelerated life test technology of DC/DC converter [J]. Environmental Technology,2019(6): 136-140.
- [15] 李磊,叶友皓,袁永生. 基于改进 Bayesian Bootstrap 方法的产品性能参数评估[J]. 电子设计工程, 2018, 26(2): 14-17,21.
- LI L, YE Y H, YUAN Y SH. Product performance parameter evaluation based on improved Bootstrap Bayesian method [J]. Electronic Design Engineering, 2018,26(2): 14-17,21.
- [16] 赵志宇,肖颖婷,任莹. 基于聚类算法的大用户用电模式识别研究[J]. 中国管理信息化,2017,20(19): 177-179.
- ZHAO ZH Y, XIAO Y T, REN Y. Research on pattern recognition of power consumption of large users based on clustering algorithm [J]. China Management Informationization, 2017,20(19):177-179.
- [17] 蔡娟,李东新. 基于优化 k 均值建模的运动目标检测算法[J]. 国外电子测量技术,2016,35(12):20-23.
- CAI J, LI D X. Moving target detection algorithm based on k-means optimized modeling[J]. Foreign Electronic Measurement Technology,2016,35(12):20-23.
- [18] 李金涛,艾萍等. 基于 k-means 聚类算法的改进[J]. 国外电子测量技术,2017,36(6): 9-13,21.
- LI J T, AI P, et al. Improvement of clustering algorithm based on k-means [J]. Foreign Electronic Measurement Technology,2017,36(6): 9-13,21.
- [19] 胡锋,朱承治,汪志华. 基于改进 k-means 算法的电力负荷分类研究[J]. 电子测量技术. 2018, 41(12): 44-48.
- HU F, ZHU CH ZH, WANG ZH H. Research on the power load classification based on improved K-means algorithm[J]. Electronic Measurement Technology, 2018, 41 (12): 44-48.
- [20] CHEN Y, XIE L, KANG R. Probabilistic modeling of solder joint thermal fatigue with Bayes method [C]. 2012 IEEE International Conference on Industrial Engineering and Engineering Management, IEEE, 2012: 787-791.

作者简介



佐磊,2013 年于湖南大学获得博士学位,现为合肥工业大学副研究员、硕士生导师,主要研究方向为智能电网技术、物联网技术。

E-mail:benzl0313@126.com

Zuo Lei received Ph. D. from Hunan University in 2013. Now he is an associate researcher and M. Sc. supervisor at Hefei University of Technology. His main research interests include smart grid technology and internet of things technology.



胡小敏,2013 年于安徽建筑大学获得学士学位,现为合肥工业大学硕士研究生,主要研究方向为电子设备的可靠性。

E-mail:2017110363@mail.hfut.edu.cn

Hu Xiaomin received her B. Sc. degree from Anhui Jianzhu University in 2013, now she is a M. Sc. candidate at Hefei University of Technology. Her main research interest includes the reliability of electronic equipment.