文章编号:1671-4598(2020)09-0272-04 DOI:10.16526/j.cnki.11—4762/tp.2020.09.054 中图分类号:TP39 文献标识码:A

样本不均衡条件下设备健康度评估方法

赵丽琴，刘 昶，邓承君
（成都大学 信息科学与工程学院，成都 610106）

摘要：为了保障设备稳定可靠运行，减少设备故障，针对很多设备采集样本不均衡的现状，提出了利用动态权重的支持向量数据描述（SVDD）方法对设备健康度进行评估；该方法首先利用设备正常健康数据进行SVDD单类学习；然后利用少量各种健康状态的数据样本计算SVDD模型超球面距离，结合其评估的健康度，使用二项式回归算法得到健康度拟合曲线，实现了健康度准确评估模型；在计算过程中，所有样本进行了指数变权的动态权重处理以提高准确性；最后以某型雷达发射机为例进行了测试验证；结果表明，该方法可实现设备健康状态准确评估，具有不错的实用价值。

关键词：支持向量数据描述；二项式回归；健康度；动态权重

Method for Evaluating Equipment Health Under Unbalanced Conditions

Zhao Liqin, Liu Chang, Deng Chengjun
(College of Information Science and Engineering, Chengdu University, Chengdu 610106, China)

Abstract: In order to ensure stable and reliable operation of equipment and reduce equipment failure, a method based on support vector data description (SVDD) is proposed to evaluate equipment health for the current situation of unbalanced equipment samples. The method firstly uses the normal health data of the device to conduct SVDD single class learning. Then using a small number of data samples of various health states to calculate the SVDD model hypersphere distance, combined with the health degree of its evaluation, using the binomial regression algorithm to obtain the fitness fitting curve, the health degree accurate evaluation model is realized. In the calculation process, all samples are added the dynamic weight processing of exponential variational weight to improve the accuracy. Finally, the test is verified by taking a radar transmitter as an example. The experimental results show that the method has good practical value for accurate assessment of equipment health status.

Keywords: support vector data description; binomial regression; health degree; dynamic weights

0 引言

健康度是设备量化的健康程度，是一种设备健康状态的定量评估，可以更准确的反映设备健康状态。随着大型机电设备系统集成化、信息化程度的提高，其故障诊断与后勤保障的难度增大。为保障这些系统连续稳定的运行，不影响任务的正常执行，减少资源浪费，提高设备保养和维修效率，需随时掌握设备的健康度，并根据系统健康度做出适当的维修维护决策，提高其工作效能。

根据文献，目前对复杂系统健康度的评估，大部分都是基于向量距离的计算，也有部分是基于模糊评判或者信息熵方法。一般常规的方法包括：综合权重法、模糊隶属度、层次分析、灰色关联、高斯模型等。如文献[1-4]就运用了综合权重和模糊评价法对设备的健康度进行了评估。文献[5]利用模糊测度和模糊积分来计算配电网关键设备的健康度。文献[6-7]利用回归算法预测健康度，取得了不错效果。也有学者用灰色关联法计算采集向量和虚拟向量的关联度，得到健康度，对设备进行健康度评估[8-9]。文献[10]分两级计算设备健康度，在部件级计算了带加权的距离，然后通过动态权重的模糊层次法计算系统级健康度。但这些方法都需要良好的先验知识来确定权重及模糊函数等信息，且不适应大量采样数据场合。

随着大数据分析技术的出现，利用机器学习方法进行健康度评估的研究越来越多。例如为了对风洞设备进行健康状态评估，文献[11]利用深度学习LSTM算法，先降维，再计算向量距离得到健康度。利用传统机器学习技术如支持向量机（SVM）算法来进行健康状态评估的研究也得到广泛重视。比如有研究采用SVM算法对高铁动车的轴承进行健康评估[12]；还有研究利用SVM算法对舰船的推进系统进行健康管理[13]。相对于深度学习算法，SVM算法占用资源少，训练样本要求不多，而准确度相差不大[14]，因此成为设备健康状态评估的主流研究方向。SVM方法是以多分类为框架，SVM通过对设备正常数据和异常数据进行学习构建分类器进行测试。但在实际测量中，容易获得大量的正常样本，而异常样本不易获得，就会造成严重的数据不平衡问题。如果异常样本匮乏，则SVM不能很好地发挥作用，对于异常情况的分类效果就不甚理想[15]。

鉴于大多数设备具备较多的正常数据，缺乏或者只有

收稿日期:2020-02-20； **修回日期**:2020-03-20。
基金项目:国家自然科学基金青年基金项目(61502059)；四川省科技计划应用基础项目(2018JY0272)。
作者简介:赵丽琴(1976-)，女，四川成都人，硕士，讲师，主要从事大数据分析、人工智能方向的研究。
刘 昶(1982-)，女，四川巴中人，博士，教授，主要从事人工智能、深度学习方向的研究。

少量异常状态数据，SVDD 作为一种单分类学习方法，成为一种评价健康度较好的备选方法。如文献［16］采用主成分分析法、SVDD 算法和马氏距离等方法，计算设备的实时健康度。文献［17］将模糊理论与 SVDD 算法相结合，提出基于模糊 SVDD 的电子装备状态评估模型。但该方法未考虑不同属性的权重，适用于样本权重均衡的场合。郑州大学的李凌均等[18]将支持向量数据描述用于机械设备状态评估研究，仅仅依靠正常运行时的数据信号，而不需要故障数据，就可以监测机器的运行状态，对早期故障诊断提供很好的帮助。

基于 SVDD 的单类学习方法在机械设备异常检测、设备故障预警，图像异常检测等领域逐渐得到广泛应用，但在设备健康状态评估方面应用甚少且准确性不高。本文以采集样本不均衡的设备为评估研究对象，提出了基于动态权重的 SVDD 设备健康度评估方法。该方法首先采用 SVDD 算法训练正常健康状态的样本得到一个超球面，然后利用少量各种健康度的标记样本计算其到超球面距离，再采用二项式回归学习算法得到健康度计算的拟合曲线。为了提高评估的准确性，本方法特别考虑了设备采样参数值在偏离最佳值越大时，健康度越差的事实，提出了基于指数函数的动态权重算法。将该算法与前面方法结合可显著提高准确性。最后以某型雷达发射机为例进行测试验证。实验结果表明，该方法对设备健康度准确评估具有不错实用价值。

1　SVDD 模型

作为一种典型的单分类器，采用 SVDD 模型，对健康数据进行训练后就可以对健康状态和非健康状态进行分类，适合缺乏全面样本的场合。

假定健康样本数据为集合 $\{x_i\}, x_i \in \boldsymbol{R}^d, i=1,\cdots,N$，训练 SVDD 模型的目标是找到一个最小体积的超球体，使所有向量 x_i 都包含在该球体内。用圆心 a 和半径 R 来表示这个超球体。为了减少奇异点的影响，引入松弛变量 ξ_i，问题转化为求解下面的二次规划问题[14]：

$$\min R^2 + C\sum_{i=1}^{N}\xi i \tag{1}$$

$$s.t.\ \ \|\varphi(xi)-\alpha\|^2 \leqslant R^2+\xi i$$

式中，C 为惩罚系数，$\xi_i \geqslant 0$（$i=1$，…，N），$\|\varphi(x_i)-a\|^2$ 为点 $\varphi(x_i)$ 到球心 a 的距离。通过建立拉格朗日函数，并引入核函数 $\kappa(x_i, x_j)=\varphi(x_i).\varphi(x_j)$ 代替内积运算，将上述二次规划问题转化为如下对偶问题：

$$\min\sum_{i=1}^{N}a_i k(x_i,x_j)-\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j k(x_i,x_j) \tag{2}$$

$$s.t.\ \ \sum_{i=1}^{N}\alpha i = 1, 0 \leqslant \alpha i \leqslant C$$

解上述不等式，得到球心：

$$\alpha = \sum_{i=1}^{N}\alpha i\varphi(xi) \tag{3}$$

取特征空间内任一支持向量 $\varphi(x_k)$，可得半径：

$$R^2 = k(xk,xk)-2\sum_{i=1}^{N}\alpha ik(xk,xk)+\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha i\alpha jk(xi,xj) \tag{4}$$

相应地，一个样本点 z 到球心 a 欧式距离 d 为：

$$d^2 = k(z,z)-2\sum_{i=1}^{N}\alpha ik(xi,z)+\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha i\alpha jk(xi,xj) \tag{5}$$

2　融合动态权重归一化算法

2.1　基于指数函数的动态权重算法

通常一个设备有多个与健康状态相关的参数，不同参数对设备健康状态的影响不同。采用权重法是一种惯常做法。在工作中发现，一个参数对健康状态的影响并不是一成不变的，随着其参数值大小变化而变化。比如某个参数平时在正常范围内，对健康状态影响不大。但是随着参数值偏离正常值越多，对健康状态的影响就越大。根据长期观察和实验验证，参数值变化对健康状态的影响程度比较贴近指数函数。因此，对于参数 i 的动态权重指数，可用如下指数函数表示：

$$w_{id} = \varphi_i^{g_i} \tag{6}$$

其中：$g_i = \dfrac{|x_i-\eta_i|}{|bound_i-\eta_i|}$

在公式（6）中 g_i 是归一化之后的参数检测值。ψ_i 是参数 i 的动态权重的指数函数参数，要求 $\psi_i>1$，其具体值根据参数在报警区域对整个设备的健康度影响程度而定。影响越大，可设定 ψ_i 的值越大。不同采集参数具有不同的 ψ 值，说明参数值增加时，对整个健康度的影响程度。x_i 是采集参数 i 的值，η_i 是参数 i 最佳值。$bound_i$ 是其边界阈值。很明显，当采集参数值超过边界值时，会快速增加，凸显该参数对整个健康状态的影响显著增强。

参数 i 的综合权重指数 w'_i 是动态权重指数 w_{id} 和静态权重指数 w_{is} 之和，如公式（7）。

$$w'_i = w_{id} + w_{is} \tag{7}$$

通常把采集的 n 个参数看做一个向量，还需要对各参数权重指数进行归一化处理，令其满足 $\sum_{i=1}^{n}w_i = 1$ 故：

$$w_i = \frac{w'_i}{\sum_{i=1}^{n}w'_i} \tag{8}$$

2.2　基于权重因子的数据归一化算法

为了得到含有权重因子的待处理数据，在数据进行归一化处理时，需要融合权重因子的影响，这样就更好地反映了各个参数对健康度的影响。含有权重的归一化采集数据值为：

$$\overline{x_i} = \frac{w_i x_i - \mu_i}{\delta_i} \tag{9}$$

在公式（9）中，x_i 是采集参数 i 的数据值，w_i 是其对应的权重因子。μ_i 则是该参数最佳值，δ_i 是其标准差，考虑计算标准差的特殊性，可采用迭代算法[19]。这样计算的结果代替原始采集数据 x_i 用于 SVDD 建模及后续 SVDD 距

离计算，可以很好地体现某些恶化参数对健康度的影响。

2.3 多项式回归拟合算法

SVDD 模型超球面反映了健康数据的范围。对每个采集数据向量 z，可以计算其到超球面圆心的距离 d，然后根据这个距离大小可以得到量化的健康状态，即健康度。通常 d 越大表示数据所代表的健康状态越差，如图 1 所示。

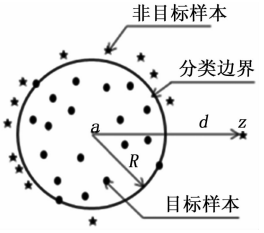

图 1 不同样本与 SVDD 超球体位置关系

通过拟合运算可以从距离 d 得到健康度。要进行拟合运算需要有各种健康状态的样本数据。但在样本不均衡情况下，很难有全面的样本数据。通常对缺乏的数据样本通过人工经验进行构造。训练样本的距离 d 和健康度 E 之间属于一种函数关系，根据设备健康度的变化特点，采用二项式回归模型具有较好的拟合效果。二项式回归的本质是通过学习，构造一条二项式曲线：

$$E_\theta(d) = \theta_0 + \theta_1 d + \theta_2 d^2 \tag{10}$$

二项式曲线应该尽可能拟合所有用于训练样本的距离 d 及其评估健康度 E。当未来的采集数据计算出距离 d 后，可以很快根据曲线函数得到对应的健康度 E_θ。

2.4 数据预处理算法

1）主成分分析法：

为了降低处理强度，在采集数据较多时，需要进行降维处理。即在尽可能保留原始数据特征的同时，尽可能降低处理数据的维度。本方法采用了 PCA 主成分分析法（Principal Component Analysis）来进行降维处理。它是一种常用的高维数据降维方法。其基本处理流程如图 2 所示。

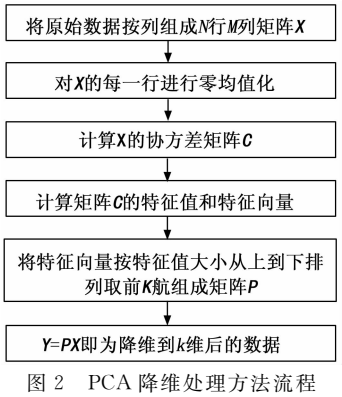

图 2 PCA 降维处理方法流程

图 2 是 M 条 N 维数据进行降维处理后，得到 k 维数据的处理过程。由于其处理过程比较简单固定，在此不做过多分析。

2）异常值剔除算法：

异常值一般是由于采集器件出现漂移或者故障出现的，对健康状态的评估具有错误的指示，需要剔除。基本思想是规定一个置信限度，凡是误差超过该限度的值认为是异常值。本文采用一阶差分法，即用两个测量值来预估新的测量值，然后与实际测量值进行比较，如果大于设定的阈值，则认为是异常值需要剔除。令 x_n 是采集值，则：

$$x'n = xn - 1 + (xn - 1 - xn - 2) \tag{11}$$

$$| xn - x'n | < w \tag{12}$$

公式（11）计算出参数的当前估计值 x'_n，在公式（12）中与真实的参数进行比较。w 为设定的阈值，与参数的变化幅度有关。

3）参数平滑处理方法：

设备检测参数由于设备的原因，免不了有噪声影响。采用平滑处理可以减少噪声影响，还可以表现参数数据的周期趋势。采用指数加权平均算法，运算量少，且具有不错的效果。

$$v_t = \beta v_{t-1} + (1-\beta) x_t \tag{13}$$

其中：v_t 是要代替的估计值，即 t 时刻的指数加权平均值。x_t 是 t 时刻采集的参数值；β 是一个权重参数（$0 < \beta < 1$）。β 越小，噪声越多，虽然可以很快适应参数的变化，但是容易出现异常值；β 越大，得到的结果越平滑，但是对参数变化的适应慢。一般需要根据参数的实际情况进行调节，得到最佳效果。一般令 $\beta = 0.9$。

2.5 健康度评估方法

综合以上算法，可以总结出基于 SVDD 的健康度评估过程如图 3 所示。整个过程分为两阶段进行：

1）在学习阶段，主要针对样本训练集进行处理。鉴于样本数据已经经过选择，一般不需要再进行预处理。这里主要根据样本向量中的参数值计算其动态权重指数，得到各个参数的权重因子，然后利用公式（9）计算含权重因子的归一化数据。如果样本处于健康状态，则进行 SVDD 超球面训练得到 SVDD 模型。如果是非健康数据，则利用 SVDD 模型计算到其到超球面圆心距离 d，根据其评估的健康度 E 进行二项式回归学习，得到计算健康度的二项式回归模型。

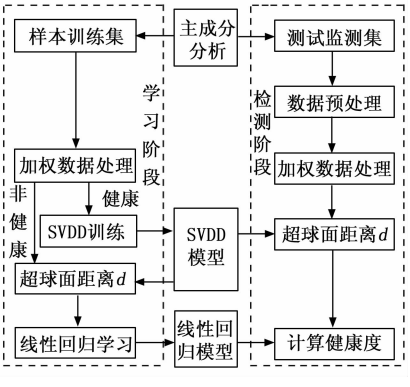

图 3 基于 SVDD 的健康度评估算法

2）在检测阶段，主要针对测试和检测集。首先进行数据预处理和权重因子计算，得到包含权重因子的归一化数

据，然后利用 SVDD 模型计算到超球面圆心距离 d，接着利用前面学习得到的二项式回归模型计算健康度，从而得到评估结果。

无论在训练阶段还是检测阶段，都利用了主成分分析的结果来选择样本向量参数，降低分析向量维度，提高分析效率。

3　健康度实例评估与分析

为了验证前述方法的正确性，利用某型雷达发射机作为实际例子进行评估分析。雷达发射机是雷达最重要的关键子系统，也是容易出现故障的部分，是重点健康管理监控设备。

1）采集数据分析，构建专家知识表：

雷达发射机的作用是在定时信号的激励下，产生大功率的射频信号。其状态监控参数较多，主要包括电磁信号参数、机械性能参数、电力参数及热参数等。为了对本文所提方法进行评估，针对某型气象雷达的发射机系统进行了实例评估。该雷达发射机采集了将近二十多个数据，但并不是所有数据都与健康状态密切有关。利用 PCA 算法及专家评估后，将与健康状态有主要影响的参数分辨出来。最后选择了 11 个与健康状态密切相关的参数作为处理数据集，得到的专家知识表如表 1 所示。表中包括各个参数的最佳值、最大值、最小值边界，以及静态权重 ws、动态权重参数 φ 等。需要根据采集值和最佳值的大小来决定 bound 值是选择最大值还是最小值。雷达专家在长期维修过程中，对这些参数的含义具有非常深刻的理解，因此填写表 1 的信息并不困难。虽然不是特别准确，但对用于验证本方法已经具有足够的准确性。

表 1　主要监测参数基本信息表

参数名称	单位	最佳值	最小值	最大值	动态权重参数 ψ	静态权重指数 ws
钛泵电流	A	0	0	20	27	5
灯丝电流	A	6.5	5	7.5	25	5
偏磁电流	A	5	2	8	50	4
注电流	mA	40	30	45	8	4
反峰电流	mA	1	0	55	10	4
整流电压	V	510	480	520	4	2
高功率电流	A	3	2	6	80	3
峰值功率	kW	270	240	300	28	2
平均功率	kW	300	260	340	24	2
发射机温度	℃	30	0	60	24	1
风道温度	℃	30	0	60	24	1

2）构建样本数据集：

在实际采集的雷达发射机监测数据中，共选取了 3 000 组数据作为训练和测试样本。其中 2 900 组数据是健康数据样本用来训练 SVDD 模型，另外 100 组数据是各种健康状态下的数据样本。选择了其中 92 组用于二项式回归学习建模，另外 8 组具有各种健康状态的样本作为测试集。由于非健康数据样本积累较少，大部分非健康数据样本利用了平时故障模型，并根据专家经验评估其健康度，作为有标记样本。在整个处理过程中，所有样本数据都需要计算动态权重，最后按照公式（9）进行加权的归一化计算，便于后续分析。

3）利用 SVDD 模型进行训练得到超球面：

首先利用其中的 2 900 组健康数据样本训练 SVDD 模型。为了比较不同情况下训练的结果，分别用两种情况下的数据样本进行了训练，即①原始数据；②利用权重因子加权的数据。两种情况下得到的超球面数据如表 2 所示。

由于是 11 维的数据向量，其圆心也是由 11 个数据构成的数组。很明显，在同样的样本下，加权处理后的数据超球面半径小很多，证明数据收敛程度更好。

表 2 不同处理情况下的超球面数据

样本处理	超球面圆心	超球面半径
未加权	[−0.205，0.057，0.141，−0.047，0.892，0.183，−0.005，−0.404，−0.302，−0.406，−0.355]	8.08
加权	[0.013，−0.448，−0.04，−0.138，−0.129，0.242，−0.079，−0.065，0.031，−0.052，−0.056]	1.10

4）利用二项式回归训练方法建立健康度拟合曲线：

使用了 92 组数据进行二项式回归训练。每个健康状态有 20 多个样本，得到样本到超球面圆心的距离与健康度的映射关系。为了对比，也是针对 SVDD 的两种情况进行分别训练，结果如表 3 所示。

表 3　二项式回归拟合结果

样本处理	θ_0	θ_1	θ_2
未加权	81.6	−3.52	0.04
加权	72.3	−1.12	0.01

5）结果分析：

8 个用于测试的样本中，每个健康状态有两个样本。与前面一样，也是分别针对两种数据处理情况，进行健康度评估测试。测试结果对比曲线如图 4 所示。其中的标定健康度是根据专家经验人工标定的结果。

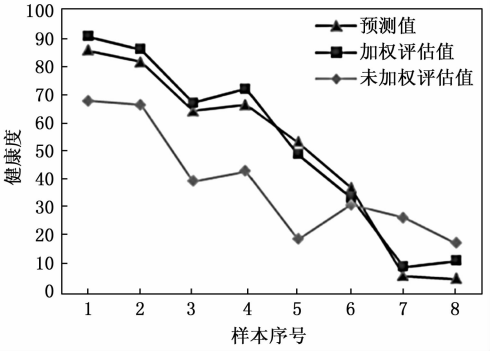

图 4　测试样本的结果分析

（下转第 281 页）