

# 导弹装备测试数据异常域检测方法研究

冯尚聪<sup>1</sup>, 杨保华<sup>1</sup>, 黄冬<sup>2</sup>

(1. 北京机电工程研究所, 北京 100074; 2. 奔牛机场, 江苏常州 213000)

**摘要:** 随着导弹装备信息化水平的快速提高, 对作战与保障的智能化应用要求不断增强, 因此, 如何快速处理导弹装备在长期使用、贮存、维护过程中积累的数据信息, 提高数据质量, 是数据时代背景下首要研究内容, 也是数据智能化应用的根本途径; 文章以大数据技术为基础, 以统计理论、深度学习等方法手段展开测试数据异常域检测方法分析, 将测试数据异常域划分为异常点、异常曲线以及异常簇, 围绕导弹装备测试数据及数据特点, 详细阐述不同特征数据条件下的测试数据异常域检测方法原理、检测模型、算法步骤等内容, 并模拟导弹装备典型测试数据序列开展应用技术研究。

**关键词:** 导弹装备; 测试数据; 异常域; 检测模型

## Research on Anomaly Domain Detection Method of Missile Equipment Test Data

FENG Shangcong<sup>1</sup>, YANG Baohua<sup>1</sup>, HUANG Dong<sup>2</sup>

(1. Beijing Electro-Mechanical Engineering Institute, Beijing 100074, China;  
2. Changzhou Airport, Changzhou 213000, China)

**Abstract:** with the rapid improvement of the information level of Missile equipment, the requirements for intelligent application of operation and support are increasing. Therefore, how to quickly process the data information accumulated by Missile equipment in the processes of long-term use, storage and maintenance, and improving the data quality is the primary research content under the background of data time and the fundamental way of data intelligent application. Based on the big data technology, the detection method of test data anomaly domain is analyzed by means of statistical theory and deep learning method, the test data anomaly domain is divided into anomaly points, anomaly curves and anomaly clusters, and by the test data and data characteristics of Missile equipment, the detection method principle, detection model, algorithm steps of test data anomaly domain are elaborated under different characteristic data conditions, and the typical test data sequence of equipment to carry out application technology research is simulated.

**Keywords:** missile equipment; test data; anomaly domain; detection model

## 0 引言

导弹装备研制、生产、使用等寿命周期内产生、积累大量的数据信息, 这些数据信息是工程应用的基础, 能够为机器学习模型提供可靠的数据集, 对于装备故障预测、质量判别、寿命预测等技术研究至关重要。由于在数据采集过程中测量精度、测量方法等因素引起的误差、数据变异、错误信息等, 导致数据信息含有不同程度的异常数据, 为数据精准应用带来困扰; 数据预处理包括数据清理、数据集成、变换以及数据规约 3 个部分, 其中过滤掉奇异点、错误干扰项、噪音数据、不一致数据等, 获得高质量的数据是数据清理的一项重要任务。针对导弹测试数据的预处理研究, 空军勤务学院王茜教授基于某型导弹测试流程, 提出了测试数据奇异项和趋势项的方法<sup>[1]</sup>, 设计数据预处理模块并进行了仿真验证。海军航空大学周璐等人为了研究基于测试数据的导弹质量评估工作<sup>[2]</sup>, 将测试数据进行筛选并调整测试模块数据权重, 最后进行归一化处理。中

国船舶工业综合技术经济研究院吴立金等人研究基于深度学习的装备故障预测技术<sup>[3]</sup>时, 对数据处理提出通过故障数据特点分析进行数据的参数化研究, 形成训练样本和识别样本, 从而删除原始数据中的伪信息。

武器系统的测试数据异常域包括异常点、异常曲线和异常簇等, 当前对异常数据的常见识别与处理方法<sup>[4]</sup>主要包括统计模型方法、距离/密度检测方法、聚类方法等, 对异常数据点的检测通常采用均方差、盒图模型、曲线拟合等统计模型的方法, 对异常数据曲线的检测采用距离模型、深度学习模型等方法, 对异常数据簇的检测采用 ITree (隔离树)、聚类模型等深度学习方法。对于数据奇异点的算法研究, 燕山大学张宁等人分别对基于统计、距离、密度、偏移等离群点检测算法进行了归纳综述, 分析了其中的优缺点<sup>[5]</sup>。晋中学院信息技术与工程学院的李俊丽等人研究了高维数据的离群点检测算法, 描述了不确定数据和数据流离群检测算法, 提出了离群组合技术用于解决高维数据的关联性<sup>[6]</sup>。广州珠江职业技术学院黄富平等人对现有的

收稿日期: 2021-11-18; 修回日期: 2022-01-04。

作者简介: 冯尚聪(1987-), 男, 山东菏泽人, 硕士, 工程师, 主要从事武器装备综合保障技术方向的研究。

引用格式: 冯尚聪, 杨保华, 黄冬. 导弹装备测试数据异常域检测方法研究[J]. 计算机测量与控制, 2022, 30(5): 127-131, 149.

异常数据检测逻辑进行改进, 利用 Hadoop 平台作为异常数据检测组的运行平台<sup>[7]</sup>, 提出利用动态反馈模组对运算后输出结果数据进行数据异常检测, 能够对隐性异常数据进行深度检测。本文针对装备数据异常域检测方法能否适用于导弹装备的测试数据进行详细探讨, 充分吸收各算法优势, 围绕导弹测试数据的特点分类提出异常域检测模型, 满足导弹测试数据在多场景工程应用中的需求。

## 1 数据异常域检测方法分析

### 1.1 基于统计模型的异常检测方法

基于统计的方法<sup>[8]</sup>假设数据符合某种特定分布, 通过数据一致性检验、包络分析确定异常点, 主要方法有符合正态分布的标准差法 (3 $\sigma$  原则)、盒图模型, 符合线性回归的最小二乘法等。标准差法即如果一个数据分布近似正态, 约 68% 的数据值会在均值的一个标准差范围内, 约 95% 会在两个标准差范围内, 约 99.7% 会在 3 个标准差范围内, 因此, 如果有数据点超过标准差的 3 倍, 那么该数据点通常判定为异常点, 如图 1 所示。盒图模型即数据通过其四分位数形成的图形化描述, 是一种简单有效的可视化异常点的方法; 四分位间距 (IQR) 将数据分为 3 个点和 4 个区间, 表征第 3 个四分位数和第一个四分位数的差 ( $IQR = Q_3 - Q_1$ ), 异常点被定义为低于箱形图下限 (或  $Q_1 - 1.5 \times IQR$ ) 或高于箱形图上限 (或  $Q_3 + 1.5 \times IQR$ ) 的观测值, 考虑把上下限作为数据分布的边界, 任何高于上限或低于下限的数据点判定为异常点; 任意一组数据集或观测值都可以根据数据的值以及它们与整个数据集的比较情况被划分为 4 个确定的间隔, 如图 2 所示。最小二乘法<sup>[9]</sup>将数据集或观测数据拟合成曲线, 根据最小二乘原理, 将实际值与利用曲线拟合计算值的离差的平方和最小为优化判据, 回归的关联式不可能全部通过每个回归数据点, 借助相关系数 “R”, 剩余标准偏差 “S” 进行判断; 将 “R” 越偏离于 1 且 “S” 越偏离于 0 的值判定为异常点, 算法原理如图 3 所示。

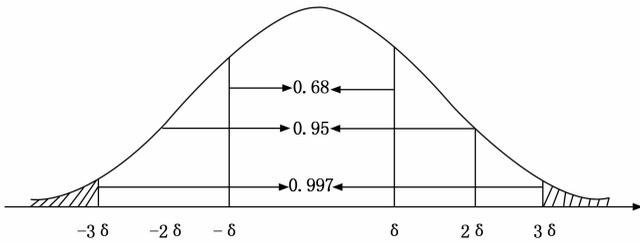


图 1 标准差法异常点检测原理图

### 1.2 基于密度的异常检测方法

相较于基于统计模型的异常点检测方法, 基于密度的异常检测方法<sup>[10]</sup>在工程应用中更常见, 通过计算异常点在数据集中的偏离度, 确定邻域密度, 度量异常点与整个数据集的临近程度, 即相似性度量或邻近性度量, 异常点临近性度量的计算方法可由到它的 k-最近邻的距离计算, 距离计算方式根据实际情况可选择欧式距离、曼哈顿距离等

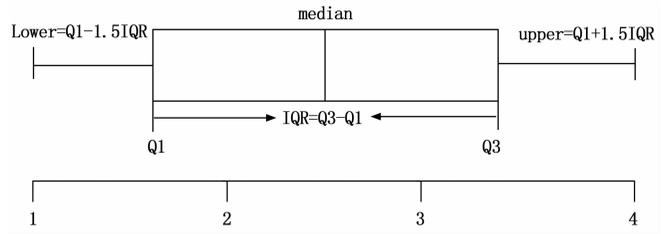


图 2 盒图模型异常点检测原理图

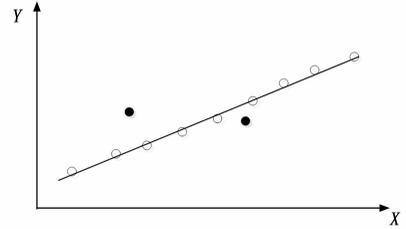


图 3 最小二乘法异常点检测原理图

距离计算公式, 临近性度量比对  $k$  的取值较为依赖, 如果  $k$  的取值太小, 则少量的相邻异常点容易计算出较低的临近性度量值; 如果  $k$  的取值太大, 而数据集点数偏少, 则  $k$  的簇中全部的数据点都有可能成为异常点, 在工程实施过程中, 通常取  $k$  个最近邻的平均距离作为邻近性度量的基准, 在计算过程中能够规避上述问题。基于密度的异常检测方法核心思想可以理解为将距离偏离度映射至异常点的密度分布区域, 即异常点在低密度区域的分布, 因此基于密度的异常点检测与邻近度的关系密切相关, 通常定义邻域密度为到  $k$  个最近邻的平均距离的倒数, 如果该距离小, 则邻域密度高, 偏离度亦小。

基于密度的异常点检测方法在定义密度概念的时候可根据数据属性进行变换, 可定义为一个数据对象周围的密度等于该数据对象指定距离  $d$  内对象的个数, 该方法依赖于  $d$  值的选取, 如果  $d$  取值太小, 则许多正常数据对象可能具有较低密度, 从而具有很高的异常点度量值; 如果  $d$  取值太大, 则许多异常点可能具有与正常数据对象类似的密度值, 因此,  $d$  值的选取影响异常点检测的准确度; 为了正确的识别数据集对象中的异常点, 需要充分利用数据集邻域相关的密度概念, 通过定义点  $x$  的密度与它的最近邻  $y$  的平均密度比值作为相对密度进行度量邻域异常状态, 使用相对密度的异常点检测步骤如下。

步骤 1: 对于指定的近邻个数  $k$ , 基于数据集的最近邻计算对象的密度  $d(x, k)$ ;

步骤 2: 依据  $d(x, k)$  计算每个对象的异常点度量值;

步骤 3: 计算点的邻近平均密度和平均相对密度。相对密度指示点是否在比它的近邻更稠密或更稀疏的邻域内, 并取作点的异常点度量比。

基于密度的检测方法最具代表性的是局部离群因子检测方法<sup>[11]</sup> (LOF, local outlier factor), 在 LOF 方法中, 为每个数据点都分配一个依赖于邻域密度的离群因子 LOF,

判断该数据点是否为离群点。判别规则为若 LOF 远大于 1, 则该数据点为异常点; 若 LOF 接近于 1, 则该数据点为正常数据点, LOF 值计算如公式 (1) ~ (3) 所示。

$$d_k(o, p) = \max\{d(o, p), d_k(o)\} \quad (1)$$

$$\rho_k(o) = \frac{k}{\sum_{p \in N_i} d_k(o, p)} \quad (2)$$

$$LOF_k(p) = \frac{\sum_{p \in N_i} \rho_k(p)}{k \cdot \rho_k(o)} \quad (3)$$

其中:  $K$  为距离邻域中的样本点数,  $d_k(o, p)$  表示数据点  $o$  到数据点  $p$  的第  $k$  可达距离;  $\rho_k(o)$  表示数据点  $p$  的第  $k$  局部可达密度,  $LOF_k(p)$  表示数据点  $p$  的第  $k$  局部离群因子。

### 1.3 基于隔离树模型的异常检测方法

基于隔离树 (iTree) 模型检测方法是一种通过隔离实现异常检测的方法, 算法原理<sup>[12]</sup>为用一个随机平面来分区数据空间, 一次可以生成两个数据子空间, 再继续用一个随机平面来分区每个数据子空间, 循环分区直到每个数据子空间里面只有一个数据点为止, 能够发现那些密度很高的簇被随机平面分区很多次是可以停止分区的, 但是那些密度很低的点很容易被分割停到一个子空间中, 从而实现将一个异类数据点从其他数据实例中隔离出来的效果。由于异常数据“少而不同”, 因此这些异常数据更容易被检测, 隔离树模型对异常产生明显的较短路径, 因为异常的实例越少, 导致分区数量越少, 树结构中的路径越短; 具有可区分属性值的实例容易在早期分区中被分离。因此, 当一个隔离树对某些特定点产生较短的路径长度时, 这些点存在很大的异常可能性。

隔离树的表示方法为设  $T$  为隔离树的一个节点,  $T$  要么是没有子节点的外部节点, 要么是有一个测试和两个子节点 ( $T_l, T_r$ ) 的内部节点, 一个测试由一个属性  $q$  和一个分割值  $p$  组成, 这样测试就可以将数据点划分为  $T_l$  和  $T_r$ 。如给定来自某变量分布的  $N$  个实例的数据样本  $X = \{X_1, \dots, X_N\}$ , 在构建隔离树的过程中, 通过随机选择属性  $q$  和分割值  $p$  递归地划分  $X$ , 直到满足: 树达到高度极限、路径为 1 或  $X$  中的所有数据都具有相同的值。隔离树隶属于二叉树, 其中树中的每个节点为零或两个子节点。假设所有实例都是不同的, 当一个隔离树完全增长时, 每个实例都被孤立到一个外部节点, 在这种情况下, 外部节点的数量为  $N$ , 内部节点的数量为  $N-1$ ; 全部隔离树的节点总数为  $2N-1$ ; 异常检测的任务是提供一个反映异常程度的排序, 因此, 可根据数据点的路径长度或异常分数对数据点进行排序, 并由此可以看出异常域是排在列表顶部的区域。基于隔离树模型检测异常的实现步骤<sup>[13]</sup>如下。

步骤 1: 隔离树模型训练。

①构建一个孤立树, 选择  $N$  个数据集对象作为样本集, 放入树的根节点;

②设定树的生长高度, 随机指定一个属性, 在当前节点数据集范围内, 随机产生一个切割点  $P$ ; 切割点满足在

当前节点属性数据中的最大值与最小值范围内;

③切割点  $P$  的选取生成了一个平面分割逻辑, 将当前节点数据空间切分为二个可度量空间: 把当前所选属性下小于  $P$  的点放在节点的左分支, 把大于等于  $P$  的点放在节点的右分支;

④在节点的左分支和右分支节点递归步骤②、③, 不断构造新的叶子节点, 直到叶子节点上只有一个数据点或孤立树已经生长到了所设定的高度。

步骤 2: 隔离树模型检测。

由步骤 1 获得  $t$  个隔离树, 将全部隔离树统一起来形成隔离树集, 即孤立森林 (iForest), 然后可以用生成的孤立森林来评估测试数据了。对于一组训练数据  $X$ , 我们令其遍历每一棵孤立树, 计算  $X$  在孤立树中的高度值 ( $X$  从根节点穿过隔离树的边数来衡量, 直到遍历在外部节点结束), 并可以计算得出  $X$  在每棵孤立树中的高度平均值。获得每个测试数据的高度平均值后, 设置一个阈值, 高度平均值低于此阈值的测试数据即为异常, 说明异常点在这些树中只有很短的平均高度。对于每个样本  $X$ , 需要对其综合计算每棵树的结果, 通过公式 (4) 计算测试数据的异常分数:

$$S(X, n) = 2^{-E(h(x))/c(n)} \quad (4)$$

其中:  $E(h(x))$  为隔离树集合中的高度  $h(x)$  的平均值,  $c(n)$  为给定样本数  $n$  的路径长度的平均值, 用来对样本  $x$  的路径长度  $h(x)$  进行标准化处理。由公式 (2) 可以得出当  $E(h(x)) \rightarrow c(n)$ , 异常得分接近 0.5, 认为样本中可能不存在异常点; 当  $E(h(x)) \rightarrow 0$ , 异常得分接近 1, 认为  $x$  为异常点; 以及当  $E(h(x)) \rightarrow n-1$ , 异常得分  $s$  远小于 0.5, 认为  $x$  为非异常点。

### 1.4 基于聚类的异常检测方法

聚类分析是在数据样本中发现的描述对象及其对象关系的信息, 将数据样本对象进行分类; 目的是各类别内的对象之间是相似的或相关的, 而不同类别中的对象是不同的、不相关的或相关性差异较大的。聚类性质表面类别内的相似性越大, 且类别间的差距越大, 说明聚类效果越好。基于聚类的异常检测方法目的是借助聚类或距离计算, 将远离所有簇中心的样本作为异常点, 将密度极低处的样本作为异常点, 基于聚类的隔离特性, 可以快速检测出异常点。本文介绍几种常用的聚类算法<sup>[14]</sup>: k-means 聚类、层次聚类、DBSCAN 聚类, 其中 k-means 聚类对于给定的样本集, 按照样本之间的距离大小, 将样本集划分为  $K$  个簇, 让簇内的点尽量紧密的连在一起, 而让簇间的距离尽量的大。层次聚类把每个样本值都视为一个类, 计算各类之间的距离, 选取最相近的两个类, 并为一个类; 新类加入进样本继续计算距离, 再合并最近的两个类, 循环归类。DBSCAN 是一种基于密度的聚类算法<sup>[15]</sup>, 通过从数据样本集中随机选择核心点, 以一个核心点为圆心, 做半径为  $r$  的圆, 选择圆内圈入点的数满足密度阈值的核心点, 且将圈内的点形成一个簇, 其中核心点直接密度可达周围的其他

实心原点，合并相互重合的簇区分类别。

设输入为样本集为  $D = \{X_1, X_2, \dots, X_m\}$ ，聚类簇数为  $k$ ，最大迭代次数  $N$ ；输出聚类簇为  $C = \{C_1, C_2, \dots, C_k\}$ ，则 k-means 聚类算法<sup>[16]</sup>步骤如下：

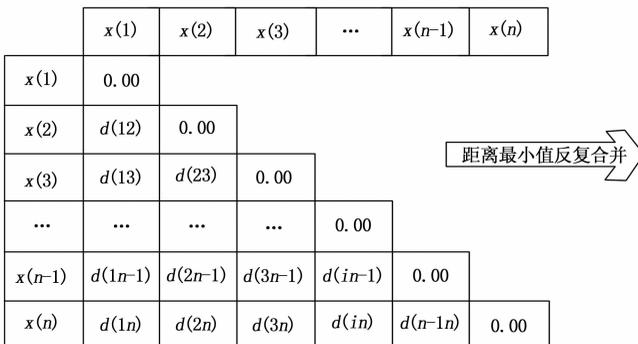
- ①从数据集  $D$  中随机选择  $k$  个样本作为初始的  $k$  个中心向量： $\mu = \{\mu_1, \mu_2, \dots, \mu_k\}$ ；
- ②初始化簇分类  $C_i = \varphi (1 \leq i \leq k)$ ；
- ③计算样本  $x_i$  和各中心向量  $\mu_j$  之间的距离： $d_{ij} = \|x_i - \mu_j\|_2^2$ ；根据距离最近的中心向量确定  $x_i$  簇标记： $\lambda_i = \operatorname{argmin}(d_{ij})$ ；将样本  $x_i$  划入簇分类，标记为  $C_{\lambda_i} = C_{\lambda_i} \cup \{x_i\}$ ；
- ④对簇分类中所有的样本点重新计算新的中心向量  $\mu_j = \frac{\sum_{x \in C_j} x}{|C_j|}$ ；
- ⑤循环步骤③、④如果中心向量不发生变化，则输出  $C = \{C_1, C_2, \dots, C_k\}$ 。

设输入为样本集为  $D = \{X_1, X_2, \dots, X_n\}$ ，聚类簇距离度量函数为  $d$ ，聚类簇数为  $k$ ；层次聚类可用树图表示聚类过程<sup>[17]</sup>，其算法流程为：

- ①计算类与类之间的距离，用邻近度矩阵表示；
- ②将距离最近的两个类合并为一个新的类；
- ③根据新的类，更新邻近度矩阵；
- ④重复步骤②、③，至剩下满足  $K$  条件的类为止；层次聚类过程如图 4 所示。

设输入为样本集为  $D = \{X_1, X_2, \dots, X_n\}$ ，邻域半径  $r$ ，邻域中数据样本数目阈值  $t$ ，邻域半径  $r$  内样本点的数量大于等于  $t$  的点叫做核心点，不属于核心点但在某个核心点的邻域内的点叫做边界点，既不是核心点也不是边界点的是噪声点。样本距离度量方式  $d$ ，可选用欧式距离计算；输出聚类簇为  $C = \{C_1, C_2, \dots, C_k\}$ ；则 DBSCAN 聚类算法<sup>[18]</sup>步骤如下：

- ①从样本集中任意选取一个数据样本点  $p$ ，计算出其它样本点到  $P$  的距离；
- ②根据每个样本点到  $p$  的距离，计算出  $p$  的  $r$  邻域；如果邻域内样本数目大于  $t$ ，则对于参数  $r$  和  $t$ ，所选取的数据样本点  $p$  为核心点，可找出所有从  $p$  密度可达的数据对象点，形成一个簇；



- ③如果选取的数据样本点  $p$  是边缘点，继续选取另一个数据样本点；
- ④重复步骤②、③，直到所有样本点被处理，输出聚类簇。

通过 k-means 聚类、层次聚类、DBSCAN 聚类等聚类方法获取数据样本的聚类簇后，根据数据特点进一步处理，将远离数据簇中心数据样本、与平均中心距离值偏差较大的数据样本或簇密度低下的数据进行异常识别和判断，以达到检测异常的目的。

### 2 装备测试数据异常域分析

装备测试数据包含系统、分系统及部件的测试数据，某种程度上能够反映装备的故障情况或健康状态，在装备使用过程中，部件或分系统的测试数据能够通过装备的分系统单元测试获取，根据不同装备的测试性设计要求，在全部的测试项目中每项测试信号异常均能表征装备的若干故障模式；当测试信号数据值在技术指标范围内时，其与技术指标标准中心值之间的偏离程度在数据量可信的情况下可以度量状态的健康状态，甚至可以预测装备的故障发生趋势。测试项目为数字量信号的，其异常值一目了然，无需过多解读，本文以测试信号模拟量为研究对象，针对装备对测试数据不同依赖需求下的异常域进行分析，将分系统或部件的测试数据异常域分为如下 3 类：

- 1) 某特征参数  $x$  在测试序列中的异常点分布情况，即  $X = \{X_1, X_2, \dots, X_n\}$ ，其中  $X_i$  表示第  $i$  次测试所产生的特征参数值，从数据序列中区分出异常点是一件较为容易的事，基于统计模型的方法、基于密度的检测方法、基于聚类的检测方法等均能检测异常点分布，其中基于统计模型的方法在测试序列异常点的检测中较为简洁且常用的方法；此需求场景多为判断分系统或部件中某特性参量在试验或长期服役后是否仍满足使用要求。
- 2) 某特征参数  $X$  在一次测试过程中随单元测试时间变化的异常曲线分布情况，即  $X = f(t)$ ，如某参数在每次测试中均满足区间内指数分布或线性分布，用  $f(t) = \omega t^{-1}$  或  $f(t) = \omega t + b$  表示，则会出现在某测试过程中随着时序变化某测试时刻出现曲线异常情况，在样本量充足的前提下基于曲线间距离偏移度、神经网络模型等均可以快速检测异常曲线，其中基于曲线间距离偏离度的方法较为常用；

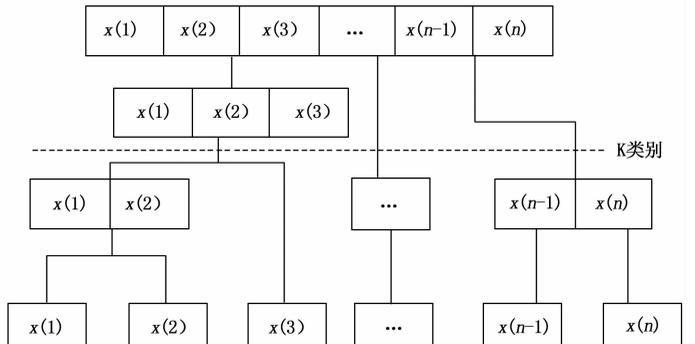


图 4 邻近度矩阵表示层次聚类过程示意图

表1 参数原始测试数据

<i>i</i>	1	2	3	4	5	6	7	8	9	10	11	12
<i>x</i>	1.247 2	1.305 9	1.479 9	1.154 8	1.526 9	1.815 7	1.347 2	1.776 8	1.442 9	1.083 2	0.247 2	0.305 9
<i>y</i>	1.805 6	1.594 1	1.560 8	1.817 0	1.278 6	1.521 5	1.345 0	1.652 1	1.607 7	1.687 6	0.805 6	0.594 1
<i>i</i>	13	14	15	16	17	18	19	20	21	22	23	24
<i>x</i>	0.479 9	0.154 8	0.526 9	0.815 6	0.347 2	0.776 8	0.442 9	0.083 2	0.079 9	2.104 2	1.546 2	1.158 5
<i>y</i>	0.560 8	0.817 0	0.278 6	0.521 5	0.345 0	0.652 1	0.607 7	0.687 6	1.482 5	1.101 0	1.417 1	1.932 4

此需求满足特定器件、电路特定参数随通电时间相关性较大的性能变化趋势分析。

3) 表征分系统某特性状态的特征参数组合, 二维及二维以上的特征向量  $s = \{x, y, \dots z\}$ , 即  $s =$

$$\begin{pmatrix} x_1 & \dots & z_1 \\ \vdots & \ddots & \vdots \\ x_n & \dots & z_n \end{pmatrix}$$

生的特征参数值, 这些特征参数之间在某种程度上又存在一定的关系, 在多次测试数据中某项测试序列与其它正常测试序列或标准测试序列存在较大偏离度或差异的情况, 认为在高维数据中存在异常簇, 基于密度的检测方法、基于隔离树的检测方法以及基于聚类的检测方法等均能检测出异常簇的分布, 针对此类问题基于聚类的检测方法应用较为广泛; 此需求场景为判断包含多个部件组成的分系统、系统的健康状态评估或故障情况分析。

根据装备测试数据特性及异常点的判别方法论述, 在工程应用中具体的实现步骤可参考如下: ①利用自动测试系统进行测试特征参数值采集; ②采集的特征参数值进行数据处理并进行在线比对分析, 直观判别超差参数; ③在参数无明显超差的情况下, 进行测试特征数据一致性分析, 寻找特征参数、参数序列之间的关系或变化规律, 确认参数是否符合某种分布; ④基于特征参数的规律特性选择合适的异常点检测方法, 评估与预测装备的状态。可结合装备的信息化管理, 建设装备参数模型库、异常域检测算法模型库, 模型库嵌入测试系统软件或装备信息管理系统软件, 将测试结果数据分别进行自动数据处理、自主数据融合、自适应模型判别、自生成辅助决策, 提升装备在大数据条件下的智能化自诊断水平。

对于复杂的导弹装备, 判断其异常域往往涉及较多分系统, 如结构、动力、制导、控制、雷达、电气等, 而表征各分系统的性能参数无论是参数数量、度量单位、技术指标均存在较大差异, 提取各分系统特征参数综合形成一个较大的特征参数满足整体导弹装备的性能需求, 特征参数构成的异常域属于高维数据, 进而演变成关于高维数据中异常点的挖掘研究<sup>[19]</sup>, 通过高维异常域的检测判别装备的性能状态存在较大的难度, 因此当前对于装备异常域的检测仍集中于分系统及部件参数, 通过各分系统的异常域的检测及分布情况, 结合分系统在导弹装备中的综合权重, 度量整个导弹装备的测试数据异常域分布, 进而评估导弹装备的健康状态或故障发生趋势。

### 3 测试数据异常域检测应用

生成一组模拟数据表示某型导弹分系统原始测试数据<sup>[20]</sup>, 如表1所示, 其中参数  $x, y$  分别为两项测试信号, 测试信号序列表征该分系统某部件当前工作状态, 从表中可以看出该数据序列每个参数项数据较为平滑, 分别应用局部离群因子检测方法及 DBSCAN 检测方法检测数据序列  $\{x, y\}$  中的异常簇, 检测结果如图5所示, 均能将数据序列异常簇检测隔离。

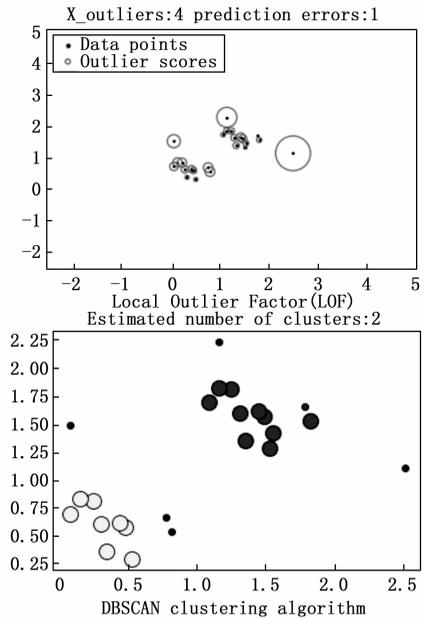


图5 局部离群因子与 DBSCAN 方法异常域检测图

### 4 结束语

本文详细描述了导弹装备在使用过程中的测试数据异常域分布及检测需求, 分析了基于统计模型、基于密度模型、基于隔离树模型以及基于聚类模型等算法检测异常域的基本原理与实现步骤, 模拟了部分算法在检测某测试序列异常簇的应用, 验证了文中所描述异常域检测方法在导弹装备测试数据处理中的应用; 但由于各类算法在收敛性、调参情况、复杂度、样本数据量需求等方面各不相同, 以及模拟数据与实际数据之间的差异性, 算法并不完全适用于多种条件下的数据异常域检测, 因此在导弹装备更多复杂数据异常域检测的应用中, 算法的实际运用效果仍需进一步探讨与工程应用。

(下转第 149 页)

旧桥梁监测中周期长、效率低、健康评估难的问题, 为桥梁的维修人员提供了参考的价值, 具有实际意义。

#### 参考文献:

- [1] 中国桥梁工程学术研究综述·2021 [J]. 中国公路学报, 2021, 34 (2): 1-97.
- [2] 李星新, 袁里, 陈湘亮, 等. 桥梁智慧检测边缘计算技术及系统开发 [J]. 湖南城市学院学报 (自然科学版), 2021, 30 (4): 1-5.
- [3] A Study on Increased Utilization of GNSS In Pier Inclination Monitoring [J]. Journal of the Korean Society for Geo-spatial Information Science, 2019, 27 (2): 43-54.
- [4] XUAN J, ZHUO Y, HUANG X D, et al. Problems in Bridge Health Monitoring and Application of Intelligent Monitoring Technology [J]. IOP Conference Series: Earth and Environmental Science, 2019, 267 (5): 1-7.
- [5] 刘康旭. 基于 Java-EE 和 Android 的桥梁健康监测系统的软件设计 [D]. 济南: 山东大学, 2019.
- [6] 张青春, 王劲尧, 何威, 等. 桥梁健康状态在线监测预警系统的设计 [J]. 仪表技术, 2019 (3): 18-20, 32.
- [7] 杜立婵, 王文静, 韦冬雪, 等. 基于 NB-IoT 的桥梁健康远程监测系统设计 [J]. 电子测量技术, 2020, 43 (20): 155-159.
- [8] 欧阳敬泓, 徐剑, 徐一超. 小桥梁的车辆荷载监测方法研究与应用 [J]. 建筑技术开发, 2020, 47 (18): 123-124.
- [9] 黄志贤. 基于 NB-IoT 的桥梁健康监测系统设计 [D].

(上接第 131 页)

#### 参考文献:

- [1] 王茜, 仲超, 杨雷. 某型导弹测试数据预处理研究与实现 [J]. 弹箭与制导学报, 2020, 40 (3): 158-160.
- [2] 周璐, 顾均元, 冯玉光. 导弹测试数据的处理与权值计算 [J]. 火力与指挥控制, 2018, 43 (12): 65-68.
- [3] 吴立金, 夏冉, 詹红燕, 等. 基于深度学习的故障预测技术研究 [J]. 计算机测量与控制, 2018, 26 (2): 9-12.
- [4] 秦艳华. 数据挖掘技术中孤立点的分析研究 [J]. 硅谷, 2010, (4): 49-50.
- [5] 张宁. 离群点检测算法研究 [J]. 桂林电子科技大学学报, 2009, 29 (1): 22-25.
- [6] 李俊丽, 芦彩林. 离群点检测算法研究 [J]. 计算机与数字工程, 2017, 45 (6): 1045-1048.
- [7] 黄富平, 梁卓浪, 邢英俊, 等. 云计算 Hadoop 平台的异常数据检测算法研究 [J]. 计算机测量与控制, 2017, 25 (7): 260-263.
- [8] 文鹤锦. 数据挖掘中统计方法应用及其分析 [J]. 中国宽带, 2021, (4): 99-101.
- [9] 白洁静. 关于插值法与最小二乘拟合的讨论 [J]. 内江科技, 2009, 30 (1): 183.
- [10] 刘冬冬. 基于密度异常因子的武器装备故障检测方法 [J]. 舰船电子工程, 2019, 39 (5): 120-123.

苏州: 苏州大学, 2020.

- [10] 张鹏飞. 基于遗传算法优化的 BP 神经网络在桥梁健康监测安全评估中的应用 [D]. 兰州: 兰州交通大学, 2021.
- [11] 申志宇. 嵌入式桥梁健康监测系统设计 [D]. 太原: 太原理工大学, 2021.
- [12] THANH Q. Nguyen. A Data-Driven Approach to Structural Health Monitoring of Bridge Structures Based on the Discrete Model and FFT-Deep Learning [J]. Journal of Vibration Engineering & Technologies, 2021: 1-23.
- [13] 邓斌. 桥梁健康监测系统在某大桥的应用 [J]. 中国科技信息, 2021 (18): 71-74.
- [14] 戴巍, 霍亚, 马尚昌, 等. Qt 下基于组件的嵌入式软件框架设计及实现 [J]. 计算机应用, 2016, 36 (S1): 257-261.
- [15] 贺伟, 刘通. 基于 Qt 的激光近炸引信信号缩比仿真软件设计 [J]. 现代电子技术, 2021, 44 (3): 45-48.
- [16] 吕焦盛. 基于 TCP/IP 协议的网络通信服务器设计与实现 [J]. 赤峰学院学报 (自然科学版), 2019, 35 (9): 50-52.
- [17] 侯彭亮, 郭莘, 王展鹏, 等. 基于 Qt 的雷达显示及手机客户端设计 [J]. 电子技术应用, 2021, 47 (8): 128-132.
- [18] 徐瑶. Qt 中基于 QCustomPlot 实现曲线绘制和显示的研究 [J]. 科技视界, 2019 (25): 54-55.
- [19] 杨学山, 侯兴民, 廖振鹏, 等. 桥梁挠度测量的一种新方法 [J]. 土木工程学报, 2002 (2): 92-96.
- [20] 刘玉玲. 考虑抗力随时间变化的桥梁可靠度评估研究 [D]. 合肥: 合肥工业大学, 2009.
- [11] 张戈, 盖赟. 局部离群因子算法 (LOF) 在异常检测中的应用研究 [J]. 网络安全技术与应用, 2020, (11): 49-50.
- [12] LIU F T, TING K M, ZHOU Z H. Isolation forest [C] // Data Mining, ICDM, 2008 8th IEEE International Conference on; Pisa, Italy, 2008: 413-422.
- [13] LIU F T, TING K M, ZHOU Z H. Isolation-Based Anomaly Detection [J]. ACM transactions on knowledge discovery from data, 2012, 6 (1): 3.1-3.39.
- [14] 孙吉贵, 刘杰, 赵连宇. 聚类算法的研究 [J]. 软件学报, 2008, 19 (1): 48-61.
- [15] 魏龙, 王勇. 基于密度划分的离群点检测算法 [J]. 计算机与现代化, 2015, (3): 26-32.
- [16] 杨俊闯, 赵超. K-Means 聚类算法研究综述 [J]. 计算机工程与应用, 2019, 55 (23): 7-14.
- [17] 乔少杰, 唐常杰, 陈瑜. 基于树编辑距离的层次聚类算法 [J]. 计算机科学与探索, 2007, 1 (3): 282-292.
- [18] 冯少荣, 肖文俊. DBSCAN 聚类算法的研究与改进 [J]. 中国矿业大学学报, 2008, 37 (1): 105-111.
- [19] 徐晓丹. 复杂高维数据中异常点挖掘算法研究 [J]. 浙江师范大学学报, 2018, 41 (2): 46-52.
- [20] 魏伟杰, 张斌, 王波, 等. 一种用于数据挖掘算法的数据生成方法 [J]. 东北大学学报 (自然科学版), 2008, 29 (3): 328-331.