

# 基于改进 BP 算法的惯性仪器故障诊断<sup>①</sup>

李富荣, 李笔锋, 王玉峰, 秦 浩

(海军航空工程学院青岛分院 仪电控制系, 青岛 266041)

**摘 要:** 为了挖掘隐藏在惯性仪器测试数据背后的信息知识, 运用数据挖掘技术, 以 Clementine12.0 为平台建立模型并实现对惯性仪器故障诊断的过程。提出一种基于两阶段聚类并做改进的 BP 算法, 与传统 BP 算法相比, 提高了预测精度和普适能力。

**关键词:** 惯性仪器; 两步聚类; k-means 聚类; 孤立点检测; BP 神经网络

## Fault Diagnosis of Inertial Apparatus Based on Improved BP Algorithm

LI Fu-Rong, LI Bi-Feng, WANG Yu-Feng, QIN Hao

(The Control of Instrument and Electronic, Naval Aeronautical and Astronautical University, Qingdao Branch, Qingdao 266041, China)

**Abstract:** In order to tap the information and knowledge hidden behind the inertial apparatus test data, applying data mining technology and taking Clementine12.0 as platform to establish model and realize the fault diagnosis of inertial instruments. Proposing BP algorithm based on two-stage clustering and its improving, compared with traditional BP algorithm, the prediction accuracy and universal capacity have been improved.

**Key words:** inertial apparatus; two-step clustering; K-means clustering; anomaly; BP Neural network

惯性仪器是陀螺仪、加速度表等惯性仪表和陀螺稳定平台以及捷联惯性测量组合等惯性测量装置的总称, 是飞行器制导与控制系统中的核心部件<sup>[1]</sup>。因此, 对惯性仪器的性能测试无论部队还是研究部门都十分重视。针对惯性仪器的测试特点, 其测试内容主要包括扩功能测试、调整测试、工艺测试、精度测试四个部分。其中, 调整测试和工艺测试是生产过程和总装测试的主要内容; 功能测试和精度测试是仪器出厂和部队测试的主要内容, 其测试项目主要包括: 电源电压测试、电流测试、电阻阻值测试、位置性能测试、回路前放测试、陀螺惯性时间测试、加速度表启动电流测试、程序功能检查、温控系统和温度状态监测等。

惯性仪器的测试是部队日常训练、执行任务及定期维护保养所必须进行的内容, 日积月累就积攒了大量惯性仪器 LRU (外场可更换单元) 的测试数据, 其中不乏典型的故障数据。面对这些海量的高维数据, 其中又有许多空缺的、冗余的、含噪声的、不一致数据, 使得传统的故障诊断方法无法直接利用这些数据

信息。另外, 由于故障的随机性、模糊性和不确定性, 一个故障的形成往往是众多因素造成的结果, 且各因素之间的联系又非常复杂, 这使得传统的故障诊断方法已不能满足现代设备的要求。而与此同时, 数据挖掘技术的迅猛发展正好可以提供一个故障诊断的应用平台, 其在特征提取、状态识别、诊断决策支持等方面具有独特优势, 在对数据处理过程中自动生成知识规则、自动学习故障诊断知识模型, 克服了专家系统知识获取瓶颈以及智能故障诊断方法所带来的诊断推理过程解释困难等问题, 为我们提供了一种如何从大量的数据中提取有效数据从而帮助我们进行正确决策分析的方法, 而基于数据挖掘技术的故障诊断也是目前一个重要研究方向。

### 1 数据挖掘过程

数据挖掘工具选用 SPSS 公司的 Clementine12.0 作为平台, 借鉴 CRISP-DM(Cross-industry Standard Process for Data Mining)行业标准, 逐步以专业理解、

① 收稿时间:2011-05-31;收到修改稿时间:2011-06-26

数据理解、数据准备、建立模型、模型评估的步骤实施惯性仪器的故障诊断过程。

1.1 专业理解

针对惯性仪器 LRU 测试数据的特点,筛选出典型故障数据,以建立一个基于数据挖掘的故障诊断分类模型作为目标。定性归纳故障字段(输入)和故障类别(输出),定量分析并建立分类模型,而通过样本学习不断完善模型,为建立基于数据挖掘技术的惯性仪器故障诊断系统提供研究基础,同时也为测试人员进行数据分析提供一种研究方法。

1.2 数据理解

待处理数据来源于两个部队同类型某飞机惯性仪器 LRU 的测试数据,时间为 2009 年 3 月至 2011 年 3 月,存储格式为 Microsoft Office Access 数据库,数据类型既有数值型也有分类型。每个惯性 LRU 都包括很多测试字段,对应相应的名称和编号,鉴于名称较长,处理时均以编号表示,下文中仅对涉及到的编号予以字段说明。

1.3 数据准备

此阶段是数据处理中的一个重要环节,约占数据挖掘过程 70%的工作量,包括数据选择、清理、重构、整合及归约等。观察分析发现,并非每条测试记录都是完整的,其中由于采集时间短或测试突然中断等原因不乏缺失值、异常值和乱码符号,处理时对于记录中缺失字段较多的予以删除,个别空缺或乱码则采用相关算法进行填充,Data Audit 节点可完成此操作。而对于文字性描述的字段需要将其数字化,例如测试通过用“0”替换,测试不合格用“1”替换,Reclassify 节点可实现此操作。

根据故障数据的分布特点,选择 8 个惯性 LRU 作为故障诊断类别,用 1~8 表示,分别为组合陀螺故障、航向位置指示器故障、角速度信号器故障、垂直陀螺故障、角位移传感器故障、垂直陀螺仪故障、角速度陀螺仪组故障和航向联系盒故障,用数字 9 表示两个以上惯性 LRU 故障,用数字 0 表示全部 LRU 工作正常。只有 LRU 的每个测试字段均合格,该 LRU 工作正常,否则即判为故障。造成每个惯性 LRU 故障的测试字段较多,不可能均用来参与建模,对于个别测试字段出现的故障记录比例甚少可认为是测试设备不稳定因素造成,例如重测正常,俗称“RTOK”问题,这样的字段应剔除。这里采用软件中的特征选择节点

Feature Selection 进行维归约,以 Pearson 卡方检验对字段的重要性进行判断,阈值取 0.65,大于阈值的为重要字段,经过筛选,每个惯性 LRU 的典型故障字段为 1~5 个,例如角速度信号器仅有的一个典型故障字段 240005 表示 A 相工作电流的起动时间测量。整理合并后形成 25 个典型故障字段作为输入,前面确定的故障类别设为输出,每个故障类别筛选出 100 条测试记录数据,其中故障类别 9 为 200 条,故障类别 0 为 500 条,共计 1500 条有效测试记录数据用于建模。鉴于数据预处理阶段每个步骤都是反复进行,数据节点流程复杂,下面仅给出数据预处理阶段针对组合陀螺的节点流程图,如图 1 所示。数据源数据依次经过过滤(Filter)、实例化(Type)、纵向合并(Append)、字段选择(Select)、属性值替换

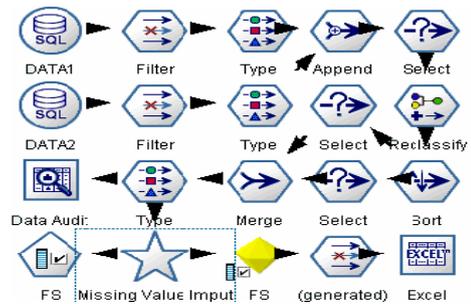


图 1 数据预处理节点流程图—组合陀螺

(Reclassify)、排序(Sort)、故障字段选择(Select)、横向合并(Merge)、空缺值补充(Data Audit)、特征选择节点(FS)后将组合陀螺的典型故障数据输出到 Eecel 表格,用于后续数据再处理。

1.4 建立模型

此次建模的目的是对惯性 LRU 故障类别做出正确的判别,判别的过程也即诊断的过程,传统的分类方法有很多,如统计算法中的 Logistic 回归,机器学习算法中的决策树,计算机算法中的神经网络等。鉴于 Clementine12.0 模块限制,选用 BP 算法作为分类方法。BP 算法是神经网络算法之一,是一种前馈式、多层、感知机网络,采用(0,1)型 Sigmoid 激活函数,非线性学习能力强,主要由训练数据对隐层的连接权值和阈值进行动态调整,以达到最佳输入输出的映射关系。

由于惯性测试中存在很多不稳定因素,诸如电源变化、环境条件、基座运动、软件漏洞、人为操作等

必然导致测试异常结果的出现,这对于分类预测意义不大,同时也会影响 BP 算法的学习能力,但需要确定这样的数据,便于测试和研究人员根据奇异值特点进行故障字段调整或 LRU 部件维修。于是提出一种基于两阶段聚类并做改进的 BP 分类算法,在两步聚类和 k-means 聚类的基础上,经过孤立点检测后再进行 BP 学习。

k-means 是使用最为广泛和经典的划分方法,它以 k 为参数,把 N 个对象分为 k 个类,使类内具有较高的相似度,而类间的相似度较低。其复杂度接近线性比较适合对大规模数据进行挖掘,伸缩性好且简单易行,具有高效性。其缺点也同样突出,一是必须指定 k 值,且对初始中心点的选择比较敏感,如果初始值选择不当,将会收敛成为一个局部最小的准则函数;二是对噪声和异常数据敏感,少量的该类数据能够对平均值产生极大的影响。而利用两步聚类算法得到的初始类可以弥补 k-means 方法对初始类选择的需求,同时通过层次聚类发现的异常值,可以通过除去异常值再将剩下的数据集作为 k-means 聚类的基础数据,有效减缓了噪声和异常数据对平均值的负面影响。在两阶段聚类基础上设置异常检测节点 (Anomaly) 来完成异常模型的建立,通过每条记录的异常指数与所设置异常指数阈值的比较来确定异常测试数据,模型生成的 \$0-Anomaly 字段给出潜在的异常记录,异常则标记为 T, \$0-AnomalyIndex 为异常指数,表示样本点所引起的类内差异与类内其他样本点所引起差异的平均值的比,该值越大,认为样本点是离群点的可能性就越大。

建立基于两阶段聚类并做改进的 BP 模型如图 2 所示。首先经过两阶段聚类 (Two Step) 得到初始类个数,即 k-means 初始聚类值 k,但为了挖掘出异常值,需要增加 k 值,通常取 1,作为异常值的类别值,通过 k-means 聚类即可得到部分异常的测试值。而经过 Anomaly 异常检测节点可进一步获得剩余数据中的少量异常值,筛选出这些异常值作为维修人员和研究人员进一步研究相关 LRU 的参考数据,剩下的测试正常数据和典型故障数据经分割节点 (Partition) 划分为训练数据和检验数据,最后由 BP 模型 (Neural Net) 进行学习和验证。每个模型节点所生成的结果都加入到数据流中,其中包含着生成权系数的 BP 模型结果节点,经过完善即可将其作为 LRU 故障诊断的分类模

型,只要将数据按照事先规定的格式存入 Excel,经过此节点即可知道数据状况以及所属故障类别。整个数据挖掘过程浑然一体,也凸显出 Clementine 软件可视化的优势。

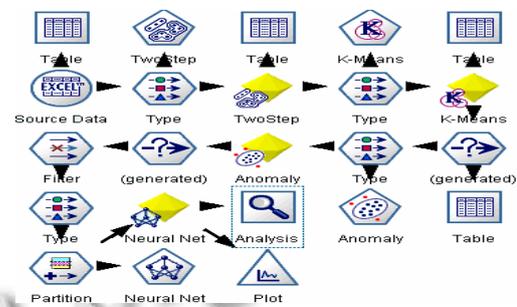


图 2 基于两阶段聚类并做改进的 BP 模型

### 1.5 结果分析及模型评价

分阶段执行模型结果,经过两步聚类最终确定两个类别,样本记录分别为 1063 和 437,简单描述为正常数据和故障数据,添加一类作为离群类的存取,即设定 k-means 聚类初始 k 为 3。k-means 聚类结果显示为三类,样本记录分别为 1335、132 和 33,可理解为正常数据和故障数据、疑似异常数据和异常数据,筛选 33 条异常测试记录,通过表格分析或散点图观察,造成异常的主要字段是 344230,字段名称为俯仰对倾斜交叉影响,是角速度陀螺仪组其中的一个字段,用于输出性能测试的描述,充分说明了角速度陀螺仪出现异常的频率较高,需要重点关注。剩余测试样本需要经过全局孤立点检测进一步确定离群点,设定最小异常指数阈值为 3,噪声水平为 0.3,噪声比率为 0.3,执行结果显示出 9 条异常测试记录数据,并且对异常影响较大的字段为 180023 和 260015,分别为航向位置指示器中的偏流同步器静子绕组电阻和垂直陀螺中的起动时间测试,而航向位置指示器和垂直陀螺作为航向姿态系统的 LRU,分别用来指示航向角和测量飞机的倾斜角、俯仰角,实际中出现故障的概率较高,也是导致飞行事故的典型仪器,需要特别关注。

在进行 BP 学习时,采用快速训练法,取 70% 样本作为训练样本,30% 样本作为检验样本,经过反复观察学习,当冲量项为 0.9,初始学习率为 0.3,学习衰减量为 30,隐单元数为 15 时分类效果最佳。利用 Clementine 的 Analysis 节点生成分类预测结果的分析值如图 3。对训练样本的分类正确率达到 97.33%,对

检验样本的分类正确率达到 98.21%。同时,建模过程可以将全部记录中影响分类程度的字段以散点图的形式显示出来,直观明了,这里任意取维组合[336020, 260015]显示分类数据的散列情况见图 4。

Results for output field 故障类别  
Comparing \$N-故障类别 with 故障类别

Partition	1_Training	2_Testing
Correct	985 97.33%	438 98.21%
Wrong	27 2.67%	8 1.79%
Total	1,012	446

图 3 BP 分类预测结果

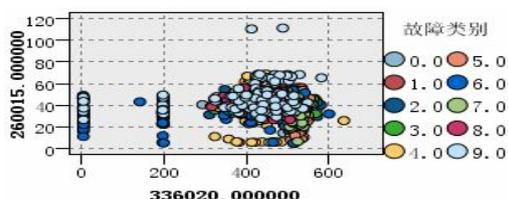


图 4 字段 336020 与 260015 分类数据散点图

利用预测准确率、查准率和查全率评估模型对惯性 LRU 故障数据的分类能力,并与传统 BP 模型做比较。预测查准率是预测为故障记录中实际故障记录的比例,体现了模型对故障记录的预测是否精确。预测查全率是实际故障记录中预测为故障记录的比例,体现了模型预测结果的覆盖程度。这三个指标越大,表明模型的预测效果越好。根据 Analysis 节点的 Coincidence matrices for \$-N 故障类别值综合计算这三

表 1 传统 BP 与改进 BP 算法预测结果指标值

算法	准确率(%)	查准率(%)	查全率(%)
传统 BP	87.47	97.95	82.62
改进 BP	97.60	98.62	97.69

(上接第 145 页)

#### 4 结语

本文提出一种区域网格方法,在常州烟草自 2010 年 5 月投入使用以来,进一步加强了对一线员工服务过程、服务绩效的考核,规范了服务过程,提升了绩效管理的水平,提升了服务质量,提升了客户服务满意度,收到满意的效果。在实际使用过程中,一线服务人员本系统也越来越理解、配合和支持,因为,本系统不仅能给他们提供有效的信息支持,也能公平地反映出他们的工作量和工作绩效。

#### 参考文献

1 王臻.突破考核的瓶颈-对客户经理绩效考核体系难点问题

个指标生成表 1,比较可知基于两阶段聚类并做改进的 BP 在分类预测效果上明显优于传统 BP 算法,可以应用于惯性仪器故障诊断之中。

#### 2 结论

本文针对大量惯性仪器 LRU 测试数据的特点,在定性分析的基础上提出一种基于两阶段聚类并做改进的 BP 分类算法,借鉴 CRISP-DM 标准实施对惯性仪器 LRU 故障诊断的数据挖掘过程,并与传统的 BP 算法做比较,提高了预测精度、普适性和泛化能力。同时对于数据挖掘技术应用于惯性测试领域具有一定的推广价值。

#### 参考文献

1 余红.惯性仪器通用测试技术应用研究.工业控制计算机,2005,18(11):19-20.  
 2 吴玉霞,牟援朝.基于两阶段聚类的洗钱行为识别.计算机工程,2010,36(15):60-62.  
 3 周生宝,郭俊芳.客户流失预测模型设计与实现.计算机系统应用,2009,18(5):170-172.  
 4 Cox JJ, Lewis RW, Ransing RS, Laszczewski H, Berni G. Application of neural computing in basic oxygen steelmaking. Journal of Materials Processing Technology, 2002,120(1-3):310-315.  
 5 Park NH, Oh SH, Lee WS. Anomaly intrusion detection by clustering transactional audit streams in a host computer. Information Sciences,2010,180(12):2375-2389.  
 6 薛薇,陈欢歌.Clementin 数据挖掘方法及应用.北京:电子工业出版社,2010.

的一点思考.上海商业,2010,7:71-72.

2 湛邵斌,陈圣波,揣媛媛,刘海博,轩义华.网格平台构建技术研究.微计算信息,2008,24(2-3):123.  
 3 宋丽,姜旭东.卷包裹法求凸包问题算法分析与程序实现.牡丹江师范学院学报,2005,4:18-19.  
 4 Xu JY, Feng YQ, Zheng ZY, Qing XH. A Concave Hull Algorithm for Scattered Data and its Applications. 2010 3rd International Congress on Image and Signal Processing (CISP2010) 2010,3: 2430-2433.  
 5 陈瑞卿,周健,虞烈.一种判断点与多边形关系的快速算法.西安交通大学报,2007,41(1):59-63.