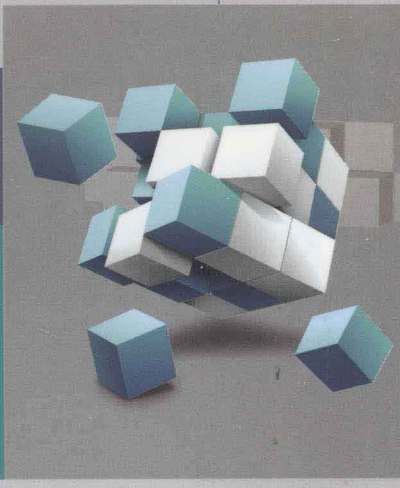
工业装备系统亚健康诊断方法

张　利　张立勇　王学芝　等著



電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

本书由国家自然科学基金（编号：61174115）资助

工业装备系统亚健康诊断方法

张　利　张立勇　王学芝　等著

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

工业装备亚健康状态预测、诊断及维护技术能够使工业装备长期稳定运行，具有重要的现实意义。本书重点介绍了几种改进及新型工业装备健康状态诊断模型，并对如何将诊断模型应用到具体问题做了详细的阐述。全书共12章，第1章综述了各种机械健康状态诊断技术的发展现状及发展趋势；第2章主要介绍了一种有效的数据预处理方法；第3~10章给出几种改进及新型状态诊断模型；第11章和第12章介绍了不完整数据集的区间重构及在此基础上的聚类算法。

本书可作为高等院校机械工程、自动化等专业高年级本科生和研究生的学习参考书，也可供工程技术人员及研究人员参考使用。

图书在版编目（CIP）数据

工业装备系统亚健康诊断方法/张利等著. —北京：电子工业出版社，2013.11
ISBN 978-7-121-21714-2
Ⅰ. ①工… Ⅱ. ①张… Ⅲ. ①工艺装备－维护 Ⅳ. ①TH16
中国版本图书馆 CIP 数据核字（2013）第 248020 号

策划编辑：秦绪军　徐蔷薇
责任编辑：王凌燕
印　　刷：三河市双峰印刷装订有限公司
装　　订：三河市双峰印刷装订有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：720×1000　1/16　印张：11.5　字数：232 千字
印　　次：2013 年 11 月第 1 次印刷
定　　价：39.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

前言

随着科学技术的发展和进步，现代化生产中的工业系统变得越来越智能化，系统的复杂性成倍增长。近年来，我国的工业化水平越来越高，在航空航天、冶金化工、核能开发等领域取得了举世瞩目的成就。而目前国内外的工业装备正向着大规模和高精度的方向发展，工业装备中各零部件间衔接的紧密程度变得越来越重要。当其中的某个或某些零部件发生故障时，可能对整台设备乃至整条生产线造成重大影响，由此会带来无法估量的经济损失，更严重的还可能造成不必要的人员伤亡。因此，装备性能的下降和意外的失效影响生产中的三个关键因素，即安全、成本和质量。人们迫切需要提高动态系统的可靠性和可维护性。其中，如何在工业装备出现故障之前，准确监测工业装备的健康状况，判断出其亚健康状态，并进行调控、诊断和维护，以使其长期且稳定运行，具有重要的现实意义。

目前，国内有关工业装备亚健康状态预测、诊断及维护技术的书籍还较为少见，出版本书意在抛砖引玉，引起国内各界对工业装备监控维护的广泛关注。本书在对基础工业装备健康状态诊断理论介绍的基础上，结合作者近几年来的研究成果，重点介绍了几种改进及新型工业装备健康状态诊断模型，并对如何将诊断模型应用到具体问题做了详细的阐述。全书共 12 章，第 1 章综述了各种机械健康状态诊断技术的发展现状及发展趋势；第 2 章主要介绍了一种有效的数据预处理方法；第 3~10 章给出几种改进及新型健康状态诊断模型，主要包括神经网络、支持向量机、小波变换和势能函数等几种方法；第 11 章和第 12 章介绍了不完整数据集的区间重构及在此基础上的聚类算法。

本书涉及的研究成果得到了国家自然科学基金项目（61174115，51104044）和辽宁省教育厅项目（L2010153）的支持。本书在出版过程中，得到了很多学者的惠助。其中特别感谢吴利春、赵家强、杨永波、郑阿楠、田立、孙丽杰、邴兆

红、蔡晓辉、张旭男、夏天、王蓓蕾、刘萌萌、张继研、罗建华、蔡连博及卢秀颖等（以出现章节为序）参与了本书部分内容撰写或提供了素材。电子工业出版社徐编辑等为提高本书质量也付出了辛勤汗水，在此一并致谢。

需要指出的是，本书的部分内容是作者所在辽宁大学计算机应用研究所及大连理工大学计算机控制研究所师生多年从事模式识别、故障诊断工作的知识积累。在此向曾经或正在研究所工作学习、在故障诊断方面做出贡献的各位老师、同学和朋友表示感谢；同时，向给予作者支持和鼓励的清华大学、东北大学等兄弟院校的老师、朋友和同行表达谢意。

由于作者理论水平有限，书中难免存在疏漏与不妥之处，恳求各位专家、学者和广大读者批评指正。

作　者

2013 年 8 月

目录

第 1 章　工业装备健康状态诊断方法论述

1.1　引言

改革开放以来，我国科学技术得到迅猛发展，工业化水平越来越高，现代化工业装备朝着大规模、高精度、复杂化过渡。装备各部件之间的联系、耦合更加紧密，当某一部件发生故障，整台设备、甚至是整条生产线都将受到影响，由此带来了无法计算的经济损失，更严重的还会产生不必要的人员伤亡。因此，复杂系统的可靠性和安全性迫切需要提高，以减少重大事故的发生。大量实践表明，健康状态诊断技术对保障设备的安全高效运行、尽早发现潜在故障、避免非计划停机、灾难性事故的发生起到至关重要的作用。因而，健康状态诊断评估技术成为一种用于解决现代化高精度系统装备可靠性、安全性的技术，是一种提高复杂设备安全使用的新方法。

对工业装备健康状态诊断方法的研究一直是一个热点，受到众多学者的关注，并取得一系列骄人的成果。清华大学信息学院的周东华教授从一个全新的视角，把现有的健康状态诊断方法分成定性分析和定量分析两大类，其中定量分析又分为基于解析模型的方法和基于数据驱动的方法，本书涉及的主要理论隶属于数据驱动方法的范畴，如图 1-1 所示[1]。本章对本书涉及的理论的基本思想和研究进展

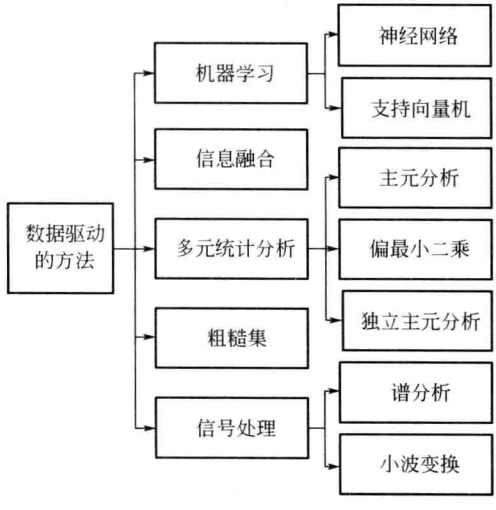


图 1-1　基于数据驱动的健康状态诊断方法分类

等做了较为详细的论述；此外，本章还概述了工业装备健康状态诊断的基本过程；最后，介绍了全书其他章节的主要内容。

1.2 粗糙集理论

粗糙集（Rough Set）是一种从数据中提炼知识并揭示暗含其中规律的数学工具，它能对各种不完备信息进行有效处理，并分析出隐藏的规律。属性约简是指在不影响整体决策分类能力的前提下，通过筛选重要或相关属性，用较少的重要属性总结出问题的正确决策。

粗糙集理论[2]的最大特点在于，它不需要任何额外的相关数据信息，仅仅依赖已存于数据内部的知识，通过数据之间的近似来表达知识的不确定性，所以它对问题的不确定性的描述或处理是比较客观的。

1.2.1 粗糙集理论的相关概念

一般地，知识被认为是比较抽象的、具有普遍意义的理论，区别于传统知识的概念，粗糙集理论中的知识被看作是一种分类能力[3~6]。

定义 1.1 由所研究对象组成的集合 U 称为论域，它是非空的、有限的。对于 U 上的一个非空子集 $\{X_1,X_2,\cdots,X_n\}$，若 $X_i \cap X_j = \varnothing$，$i \neq j$，并且 $\cup_{i=1}^{n} X_i = U$，那么这个非空子集 $\{X_1,X_2,\cdots,X_n\}$ 就是 U 上的一个划分。

定义 1.2 若集合 A 上的二元关系 R 满足自反关系、对称关系、传递关系，则称 R 是一个等价关系。对于 $\forall a \in A$，a 关于 R 的等价类 $[a]_R = \{x \mid xRa\}$，简称 $[a]_R$。用 U/R 表示，对于论域 U，在等价关系 R 下的一个划分。

定义 1.3 知识系统 S 可以表示成有序的四元组 $S = <U, A, V, f>$。其中，属性集 A 是由条件属性 C 和决策属性 D 的并集组成；$V = \cup_{r \in A} V_r$ 为属性值的集合，V_r 表示第 r 个属性的取值，$f: U \times A \to V$ 是一个映射函数，确定 U 中任意一个对象 x 的值。

定义 1.4 任意属性集 $\forall P \subseteq A$，知识系统 S 上的不可分辨关系 $\mathrm{IND}(P)$，可定义为：$\mathrm{IND}(P) = \{(x,y) \in U \times U : \forall b \in P, b(x) = b(y)\}$。

如果 $(x,y) \in \mathrm{IND}(P)$，则称 x 和 y 对 P 是不可分辨的。显然，不可分辨关系 $\mathrm{IND}(P)$ 为一等价关系，且 $\mathrm{IND}(P) = \cap_{b \in P} \mathrm{IND}(\{b\})$。$U/\mathrm{IND}(P)$ 是不可分辨关系 $\mathrm{IND}(P)$ 对论域的一个划分，记作 U/P。

定义 1.5 对于任意子集 $\forall X \subseteq U$，R 为不可分辨关系，给出 X 在知识系统 S

中关于 R 的下近似集合和上近似集合的定义：

$$\underline{R}(x)=\{Y\in U/\mathrm{IND}(P)|Y\subset X\} \tag{1.1}$$

$$\overline{R}(x)=\{Y\in U/\mathrm{IND}(P)|Y\cap X\neq\varnothing\} \tag{1.2}$$

称 $\mathrm{POS}(X)=\underline{R}(x)$ 为正域，$\mathrm{NEG}(X)=U-X$ 为负域，$\mathrm{BN}(X)=\overline{R}(x)-\underline{R}(x)$ 为边界域。

由定义 1.5 可知，在关系 R 下，由那些肯定属于 X 中的元素构成了下近似集合 $\underline{R}(x)$，又或者说是正域 $\mathrm{POS}_R(X)$；而由那些肯定不属于 X 的元素构成了负域 $\mathrm{NEG}_R(X)$；上近似集合 $\overline{R}(x)$ 是由那些肯定属于或可能属于 X 的元素组成的集合；而边界域 $\mathrm{BN}_R(X)$ 是那些不能确定的元素组成的集合。

在现实生活中，使用众多特征属性来尽可能细致地描述对象，这些属性的重要程度是不同的，甚至说某些属性是冗余的。删除这些冗余的属性和不必要的属性值，不但不会影响知识发现的结果，还能达到简化计算的作用，同时也能降低算法复杂度。

定义 1.6　设 R 是一等价关系集，且 $r\in R$，若有 $\mathrm{IND}(R)=\mathrm{IND}(R-\{r\})$ 成立，那么 r 就是不必要的；否则 r 是必要的。若对于 R 中的任何一个关系 $\forall r\in R$ 都是必要的，那么称等价关系集 R 是独立的，否则称 R 是依赖的。

定义 1.7　设 $Q\subseteq R$，若 Q 是独立的，且 $\mathrm{IND}(Q)=\mathrm{IND}(R)$，则称 Q 是等价关系族 R 的一个约简，记作 $\mathrm{Red}(R)$。

定义 1.8　在等价关系 R 中，由所有的必要关系构成的关系子集称为等价关系 R 的核，记作 $\mathrm{Core}(P)$。所有约简 $\mathrm{Red}(R)$ 的交集等于 R 的核，即 $\mathrm{Core}(P)=\cap\mathrm{Red}(R)$。

1.2.2　常用的属性约简算法

属性约简问题，作为粗糙集理论的核心，长久以来一直得到众多学者的关注，先后出现众多行之有效的约简算法。在实际应用中，没有必要获得属性的最佳约简，只要得到一个有效的相对约简就已经足够可以有效解决具体问题了。

下面详细介绍两种常用的属性约简算法[7~9]。

1．基于属性重要度的属性约简

由核属性 $\mathrm{Core}(P)$ 的定义可知，各约简结果的交集一定属于核属性，所以先找到 $\mathrm{Core}(P)$，有助于实现属性的快速约简。基于属性重要度的约简算法，就是在求得属性集的核属性的基础上，通过扩充得到属性子集的一种约简算法。

属性重要度计算公式为：

$$\mathrm{SGF}(a,\mathrm{Red},X)=(|\mathrm{POS}_{a\cup\mathrm{Red}}(D)|-|\mathrm{POS}_{\mathrm{Red}}(D)|)/|U| \tag{1.3}$$

从非核属性集中找出部分属性，然后将其与核属性合并所得并集计算对决策

属性的相对正域，当遍历完所有条件属性后，由相对正域相等的添加属性组成的集合就是一个约简结果。

具体实现步骤如下：

1） 首先求得属性集的核属性 Core

输入：信息系统 $S=(U,A,V,f)$。

输出：Core，表示核属性集。

（1）加载数据 data，并令 Core $=\varnothing$；

（2）$\forall a\in A$，如果 $\mathrm{IND}(A-\{a\})\neq \mathrm{IND}\{A\}$，则 $\mathrm{Core}\Leftarrow \mathrm{Core}(Y\{a\})$；

（3）当遍历完 A 中元素时，算法终止。

2）属性约简

输入：信息系统 $S=(U,A,V,f)$。

输出：Red，是一个属性约简的结果。

（1）输入数据 data，求出 S 的核属性集 Core；

（2）令 Red=Core，若 IND(Red)=IND(A)，结束算法，输出 Red；

（3）计算 $\forall a\in A-\mathrm{Red}$ 的属性重要度并按降序排列，选取属性重要度最大的，当属性有多个时，则选取划分个数最少的，$\mathrm{Red}\Leftarrow \mathrm{Red}(Y\{a\})$；

（4）如果 $\mathrm{IND(Red)}\neq \mathrm{IND}(A)$，转到步骤（3），否则算法终止，此时的 Red 即为所有约简属性集。

2. 基于信息熵的属性约简

我们可以把等价关系认为，由论域 U 的任意子集构成的随机变量。

定义 1.9 等价关系 R 的信息熵 $H(R)$ 可以按下式来计算：

$$H(R)=-\sum_{i=1}^{n}P(X_i)\log P(X_i) \tag{1.4}$$

当 $H(R)=0$ 时，说明只有一种可能性存在，没有其他不确定情况发生；相对于有 n 种可能发生的情况，$H(R)$ 取得越大，则系统的不确定性越大。

具体实现步骤如下：

输入：信息系统 $S=(U,A,V,f)$。

输出：Red，是一个属性约简后的结果。

（1）按公式（1.4）计算 $H(A)$ 的值；

（2）计算 S 的核属性集 Core，并在此基础上进行扩充，令 Red=Core；

（3）当 $H(\mathrm{Red})\neq H(A)$ 时，顺序执行，否则跳转到步骤（5）；

（4）对于 $\forall a\in A-\mathrm{Red}$，计算 a 的重要度：

$$H(\mathrm{Red}(Ya))-H(\mathrm{Red})=\sum_{i=1}^{m_1}P(X_i)\log_2 P(X_i)+\sum_{i=1}^{m_2}P(Y_i)\log_2 P(Y_i) \tag{1.5}$$

其中，$\{P(X_i)\}_{i=1}^{m_1}$，$\{P(Y_i)\}_{i=1}^{m_2}$ 分别是 Red($Y\{a\}$) 和 Red 的概率分布，取差值最大的属性 a，并令 Red $\Leftarrow$ Red($Y\{a\}$)，跳转到步骤（3）；

（5）算法结束，输出 Red 。

1.3　神经网络

神经网络是一种机器学习算法，而反向传播网络（简称 BP 网络）作为一种应用最广泛的神经网络，具备一套完整的理论体系和成熟的学习机制。它模仿人脑神经元对外部激励信号的反应过程，建立多层感知器模型，利用信号正向传播和误差反向调节机制，通过多次迭代学习，搭建出处理非线性信息的网络模型。BP 网络及其算法是人工神经网络的精华所在，它被广泛应用于函数逼近、模式识别、故障分类等。而在实际应用的人工神经网络，80%～90%都是直接采用 BP 网络或是其变种形式[10]。

1.3.1　BP 神经网络结构

通常，BP 神经网络由输入层、隐含层和输出层构成，相邻两层的节点彼此互相连接，但位于同一层的节点间没有任何连接。隐含层节点个数目前尚无确定的标准，可通过反复试凑的方法，以相对较好的实验结果确定节点数。根据 Kolmogor 定理，我们知道：一个三层的 BP 神经网络，如果隐含层采用 Signal 型的激活函数，那么在闭区间上，能以任意精度逼近相应的非线性连续函数。

为了简化说明，本节选择具有单隐含层的 BP 神经网络进行说明，即只有一层隐含层。如图 1-2 所示为一个具有单隐含层的 BP 神经网络模型[11]。

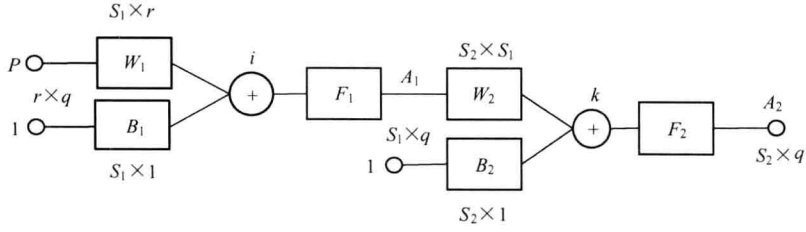


图 1-2　具有单隐含层的 BP 神经网络模型

在 BP 算法中，各层神经元节点的激活函数必须是处处可导的，因此处处可导的 S 形曲线被经常使用，如对数函数和正切函数。在实际操作中，纯线性函数常用作输出层的激活函数，如图 1-3 所示为对数函数 logsig()和纯线性函数 purelin()。

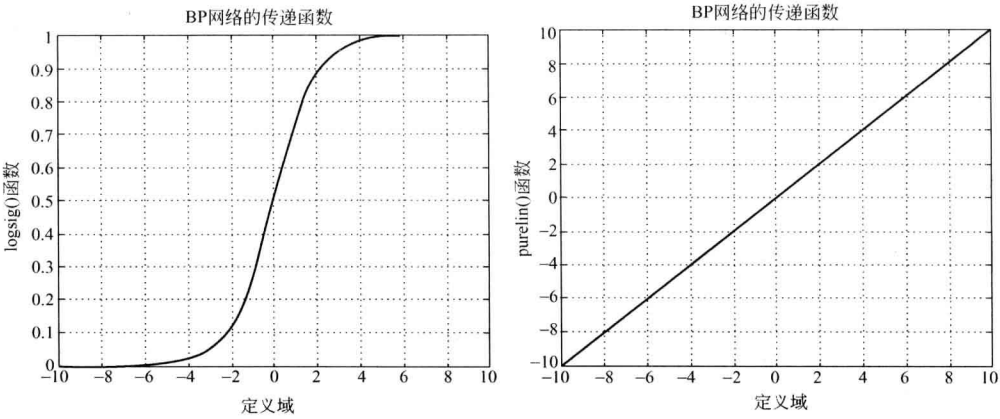


图 1-3 激活函数 logsig()和 purelin()

1.3.2 BP 算法的步骤

1. 标准 BP 算法的主要思想

标准 BP 算法[12]把整个训练过程分成两个阶段，即信号的正向传播阶段和误差的反向调节阶段。在正向传播过程中，输入信号在权值和阈值的作用下经隐含层传向输出层，并在输出端产生累加计算结果。这一过程中，存在于网络中的连接权值和阈值保持不变，并且上一层神经元的输出只能作用于与其相邻的下一层神经元节点。当正向传播的结果与样本对的期望输出间的误差不能满足预先设定的误差精度时，转到误差的反向调节阶段。在误差的反向调整过程中，误差信号由输出端开始，按照学习算法向前传播，并将误差动态地分配给连接输出层与隐含层、隐含层与输入层的神经元间的权值。经过这种反复的调节，神经元间的权值和阈值得到不断的修正，并趋于稳定。在正向传播的计算结果与样本对的期望输出间的误差满足预先设定的误差精度或达到最大学习次数时，停止训练学习。

2. 标准 BP 算法的主要步骤

为了简化说明，本节以典型的三层 BP 网络为例，详细描述 BP 算法的计算流程，如图 1-4 所示。标准 BP 算法是一种基于梯度下降的学习算法，当网络输出与期望输出的误差达不到要求时，通过误差的反向调整过程，把误差分摊到存在于网络中的权重中，使正向传播的计算结果与样本对的期望输出间的均方误差趋于最小得以实现。