

# 工业大数据环境下的混合故障诊断模型研究

孔宪光, 钟福磊, 马洪波, 李继刚

(西安电子科技大学 工业大数据技术研究中心, 陕西 西安 710071)

**摘要:** 随着传感器、物联网、大数据等技术的广泛使用, 工业领域已经正式进入工业大数据时代, 工业大数据具有明显的实时性、动态性和不确定性等特点, 产品和装备被赋予了更多的智能化和海量数据的属性, 解决了传统产品运维端采集数据量不足的问题。故障的检测和诊断一直是智能装备领域研究的重点, 它可以提高设备的可靠性、可用性以及可维护性。故障检测和诊断方法可以分为: 基于数据驱动的方法, 基于机理建模的方法, 基于知识工程的方法。基于工业大数据技术优势并符合故障诊断的可靠性要求, 本文将基于知识工程的方法和基于数据驱动的方法相结合, 建立了一种新的混合故障诊断模型。最后以盾构机为研究对象, 将混合模型应用在盾构机的健康状况管理, 提出了针对盾构机的故障诊断模型。

**关键词:** 工业大数据; 故障诊断; 基于数据驱动的方法; 基于知识工程的方法

## 0 引言

随着传感器、嵌入式终端系统、智能控制系统、通信设施、智能制造等相关技术的发展, 以及互联网、移动互联网、物联网等新一代信息技术在工业领域的应用, 工业企业所拥有的数据也日益丰富, 工业也进入大数据时代。工业大数据的主旨是通过充分利用信息通讯技术, 推动制造业向基于大数据分析与应用基础上的智能化转型。工业大数据技术能够提升制造业的效率和创新能力。辛辛那提大学的 Jay Lee 教授指出在“现代的制造业中, 存在着许多的不确定因素。他将这些不确定问题分为两部分: 可见问题和不可见问题。可见问题是同一看到并能够得到解决的。人们经常会忽略不可见问题, 例如设备的加工性能下降、零部件磨损等。人们一般会认为一台设备可以持续运行很长一段时间, 但是在真实的工厂生产中, 设备每时每刻都在损耗。工业大数据提供透明化工具和技术, 使得设备运行过程中不断产生的数据可以被系统性的处理为信息。利用这些信息来解释不确定因素, 从而使管理者可以掌握信息, 然后做出有把握的决定”<sup>[1]</sup>。

伴随着复杂大型的设备在制造业中的发展与应用,

其研制、生产、维修的成本也越来越高。同时, 由于施工环境和影响因素的增加, 故障的发生率逐渐加大。基于复杂系统可靠性、安全性的考虑, 故障预测和健康管理 (Prognostics and Health Management, PHM) 策略受到越来越多的重视和应用。PHM 是从状态监控向健康管理的转变, 主要目的提高装备系统可靠性和安全性, 从而以较小的维修成本实现基于状态的维修 (Condition-Based Maintenance, CBM)<sup>[2]</sup>。利用工业大数据分析技术, 在设备的故障将要发生的时候对设备进行维护, 这种维护是根据状态检测、故障诊断分析的结果。这是一种主动、积极的维护方式, 尤其是针对一些利用率高、发生故障后对生产影响较大的关键设备, 或配件供应困难、精度高的大型设备有必要实行基于状态的维护。

目前, 故障诊断方法可以划分为: 基于机理建模的方法、基于知识工程的方法、基于数据驱动的方法<sup>[3]</sup>。在工业大数据环境下, 当面对复杂设备时, 只能获取设备的部分信息, 而并不能获取全部信息, 难以建立精确的机理模型。基于数据驱动的方法可避免建立精确的数学模型, 所以逐渐成为研究的热点。尤其是在工业大数据环境下, 基于数据驱动的方法得到了充分的应用。

(1) 基于机理建模的方法: 充分体现设备的内部机理, 变量的物理意义明确, 有很好的外延性。但是当面对复杂设备系统时, 无法获取内部机理的全部信息, 将

教育部中央高校基本业务费大数据项目群 (7214487602)

会难以建立精确机理模型。另外，机理分析总是基于很多假设和简化，所以模型状态和输出信息将会存在一定的误差。

(2) 基于知识工程的方法：不需要建立精确的定量机理模型，而且诊断结果易于理解，适应具备大量生产经验和工艺知识的情况。但是通用性差，且需通过大量的经验知识建立知识库。当出现一种新的未知故障，由于没有故障的任何知识，容易出现误报和漏报的现象。基于知识的方法的优点是它不需要对系统进行数学建模，该方法引入检测对象的多种信息，如设备定性的结构知识、专家知识、决策知识等。

(3) 基于数据驱动的方法：不需要设备精确模型和先验知识，只需要对设备数据进行处理与分析，但是数据模型的参数不具有明确的物理意义，难以实现有效隔离和故障诊断。数据驱动方法就是利用设备的历史和实时的数据等，通过对这些数据进行处理，然后进行故障的检测等。该方法的主要优点是利用了大量的设备运行数据，采用数学方法进行有用信息的提取，为现场人员提供有用的统计数据和推理信息。

国内外一些专家在混合建模方面也有一些相应的研究，Ahmad Alzghoul 等对液压驱动系统进行故障检测和预测，将基于数据驱动方法和基于知识工程的方法进行了对比，但并没有将这两种方法结合起来进行故障的处理<sup>[9]</sup>。孙冬将基于机理建模的方法和基于数据驱动的方法结合起来进行建模，并在三容水箱系统和四旋翼飞行器控制系统中进行了应用研究<sup>[9]</sup>。刘强等人提出基于数据和知识的工业过程监控及故障诊断方法<sup>[9]</sup>，他将该混合方法应用在工业过程中。

总的来说，对于基于数据驱动方法和基于知识工程的方法结合起来进行工业设备的故障处理研究还没有成熟的体系和方法。本文针对工业大数据环境下的设备故障诊断问题，为了提高故障诊断的可靠性，提出一种混合模型故障检测模型，利用数据驱动方法进行故障检测，应用基于知识工程的方法来对故障进行诊断。

本文的首先阐述了论文的研究背景以及故障检测的各种方法的优缺点。然后介绍了混合模型总框图，以及基于数据驱动的方法和基于知识工程的方法应用的算法。

## 1 混合故障诊断建模方法

### 1.1 混合故障诊断建模框架

将基于知识工程的方法和基于数据驱动的方法进行

结合，主要利用基于数据驱动的方法来对故障进行检测，利用基于知识工程的方法来对故障进行诊断。混合建模方法框架如图 1 所示。

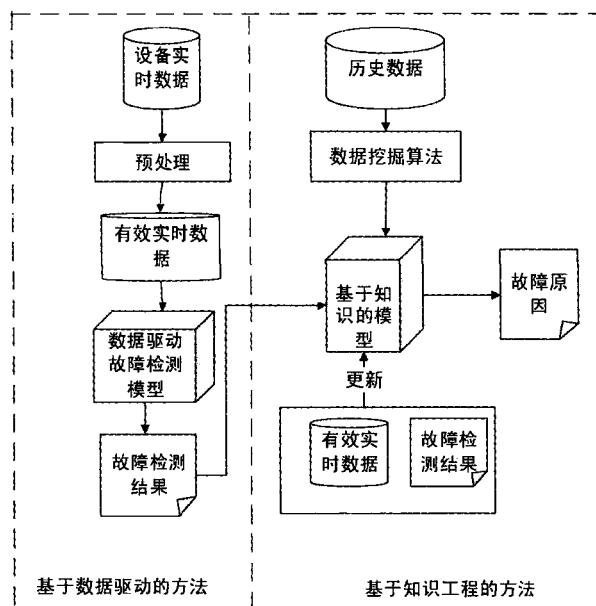


图 1 混合故障建模方法框图

实现的流程如下：

(1) 对设备的实时数据进行收集、预处理、特征提取后得到有效的数据。

(2) 利用基于数据驱动方法建立的故障检测模型对数据进行处理，得出故障检测的结果。

(3) 对历史数据进行挖掘，得出设备领域的知识和规则等，建立基于知识工程的模型如专家系统等。

(4) 将故障检测结果输入到基于知识工程的模型中，对故障进行诊断，得到故障的原因，结合知识模型实现决策支持。

(5) 若利用知识模型没有对故障诊断成功，实时数据和故障检测结果进行数据挖掘来实时的更新基于知识的模型。对实时的数据进行数据挖掘，得到出现的关键字，重要字段等，更新知识模型。

### 1.2 基于数据驱动的方法

支持向量机算法 (support vector machine) 是一种小样本学习方法，从发展到现在已经有着十分坚实的理论基础了，简化了通常的分类问题。利用支持向量机算法来建立模型，结构风险低，而且泛化能力强，可以用来对特定领域和学科进行综合评价和预测。因此，本文采用其作为数据驱动的方法。

支持向量机建立模型的流程图如图 2 所示。

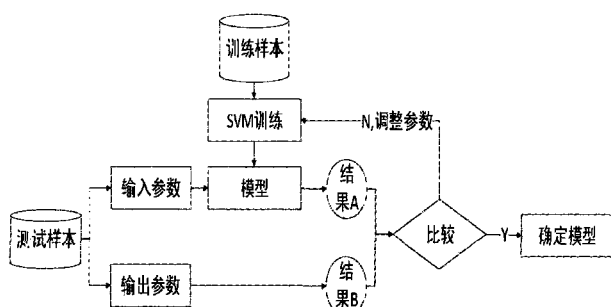


图 2 支持向量机建模流程图

将数据样本集分为测试样本和训练样本。首先利用训练样本对支持向量机模型进行训练，训练完成后将测试样本输入到训练好的模型中，检查模型的准确度是否达到规定要求，如果达到了规定要求，就确定当前模型。否则调整参数并重新训练模型。

SVM 核心是找到一个最优的分类超平面，该超平面可以将数据进行分类，最大化训练集的间隔。线性分类图如图 3 所示。将两个样本分开的实线叫做分类超平面。

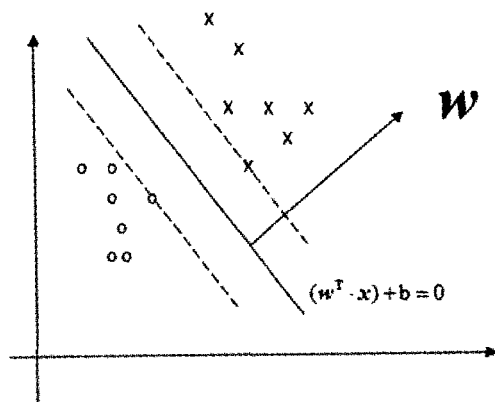


图 3 线性分类图

(1) 线性可分的情况。假设有两类数据，给定训练样本如公式 (1) 所示。

$$(x_i, y_i), i=1, 2, \dots, l, x \in R^n, y \in \{+1, -1\} \quad (1)$$

将超平面以及它的约束定义为：

$$(\omega^T \cdot x) + b = 0$$

$$s.t. y_i(\omega^T \cdot x_i + b) \geq 1, i=1, 2, \dots, n \quad (2)$$

为了能够方便快速的求解出这个最优化问题，应用 Lagrange 对偶性，为每一个约束条件加上一个 Lagrange 对偶变量  $\alpha$ ，经过一系列的变化转化为求解对偶变量  $\alpha$  的最优化问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ s.t. \quad & \alpha_i \geq 0, i=1, 2, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (3)$$

解得最优解  $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*)^T$ 。

(2) 非线性的情况。支持向量机的处理方法是选择一个核函数  $k(\cdot, \cdot)$ ，通过将数据映射到高维空间，来解决在原始空间中线性不可分的问题。我们把计算两个向量在隐式映射后的空间中的内积的函数叫做核函数 Kernel Function，利用核函数可以直接在原来的低维空间中进行计算。

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \\ s.t. \quad & \alpha_i \geq 0, i=1, 2, \dots, n \\ & \sum_{i=1}^n \alpha_i y_i = 0 \end{aligned} \quad (4)$$

其中  $k(x_i, x_j)$  是核函数。

利用核函数就很好的避开了直接在高维空间中进行计算，但是计算的结果却是等价的。

### 1.3 基于知识工程的方法

故障树分析法对故障的处理方式类似于人类的思维方式，很容易理解，技术已经相对成熟，在许多的领域都有应用。它能够对故障进行快速的判断，提高了判断的效率。因此，本文采用故障树分析法。

故障树分析法 (Fault Tree Analysis, FTA) 是一种图形演绎法，应用一些逻辑关系符号、事件符号等，来对事件之间的逻辑和因果关系进行表示。它通过将系统故障与导致系统故障的诸多因素通过故障图表的方式形象的表达，直观的反映出故障、系统、元部件、故障原因之间的关系。它是一种根据系统可能发生的事故或已经发生的事故结果，去寻找与该事故发生有关的原因、条件和规律，同时可以辨识出系统中可能导致事故发生的危险源。故障树分析法建立模型的流程图如图 4 所示。

将设备的状态输入到故障树的模型中，检查现在的模型中是否存在与输入状态相对应的规则，如果有那就利用规则来判断结果，如果没有的话，就利用其它的手段来判断设备的状态，得到结果后，更新现有的故障树模型。运行时间越长，故障树模型里面的规则就越丰富，判断的精度和实时性也就会越高。

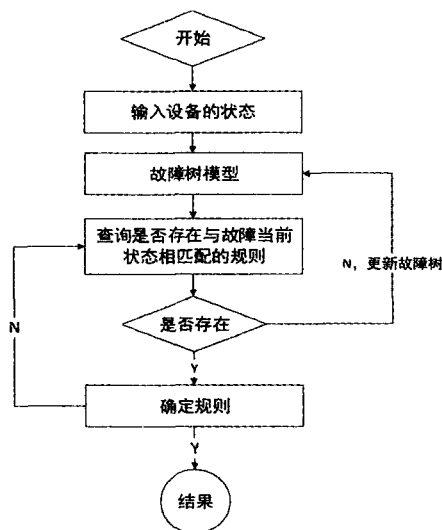


图4 故障树建模流程图

在故障树分析法中，将系统最不希望发生的故障状态称为顶事件，将导致故障状态发生的所有可能的直接原因事件称为中间事件。将导致中间故障事件发生的所有可能直接原因称为底事件，底事件不能再分解。所有可能的故障模式组合称为最小割集。故障分析法包括定性分析和定量分析。定性分析分析系统的某个故障由哪些因素引起。能够寻找出故障树的所有最小割集。定量分析的目的是为了得到底事件互相独立和已知其发生率下，顶事件发生的概率和底事件重要度等定量指标。

利用故障树还可以进行重要度分析。重要度分析就是对最小割集或部件的重要度进行计算，重要度是指一个零件、部件或最小割集对顶事件的贡献。在实际中，要根据具体情况选择。具体步骤：（1）求出支持故障检测的全部最小故障树。（2）按重要度将最小割集排列。（3）排除当前故障。

## 2 基于混合模型的盾构机故障诊断

### 2.1 总体流程图

盾构机故障诊断的主流程如图5所示。

盾构机故障诊断是盾构机健康管理的主要内容。上述流程图将大数据驱动判别和专家知识库判别相结合。结合失效模式、故障模式以及关键因素分析，综合形成故障诊断记录，作为故障解决方案的基础。

### 2.2 盾构机故障诊断专家系统结构设计

工业大数据环境下的盾构机PHM专家系统如图6所示。

与传统PHM专家系统相比。盾构机PHM专家系统

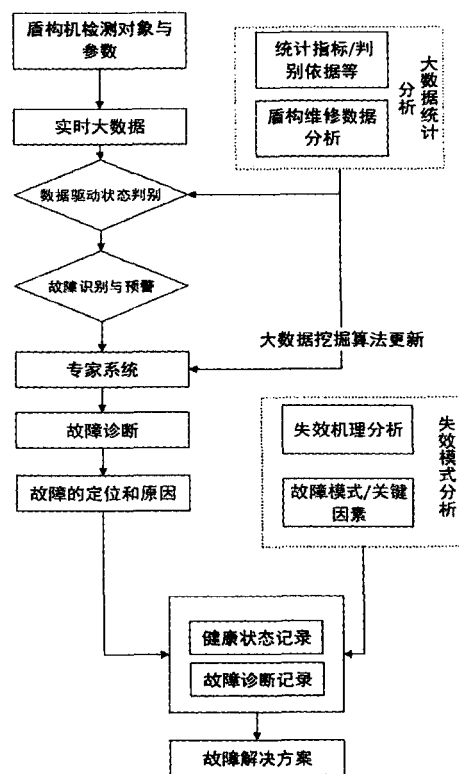


图5 盾构机故障诊断主流程图

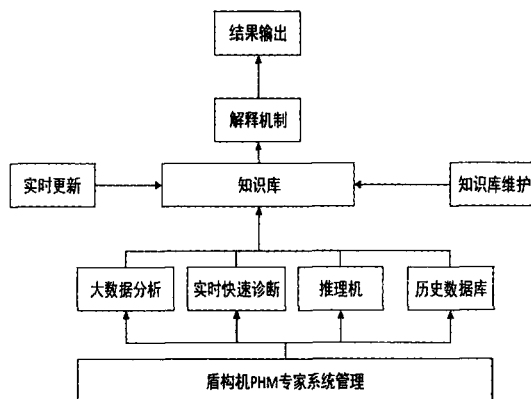


图6 基于大数据的盾构机PHM专家系统图

管理最大的不同在于大数据分析结果的知识提取和实时更新机制的引入。知识库来源于专家知识与实时大数据，这将会对系统构架、管理和维护提出新的要求。

## 3 结论

本文主要提出了一种工业大数据环境下的混合故障诊断模型，将基于知识工程的方法和基于数据驱动的方法相结合，提高设备故障检测和诊断的可靠性。首先介绍了混合模型的整体框架以及如何将两种方法有效的结合在一起。然后分别介绍了两种方法所用到的算法。最

后讲述了混合模型在盾构机中的应用。本文的主要贡献是通过将这两种模型结合起来,并将工业大数据分析技术应用在盾构机的故障诊断中,为故障处理和可靠性方向提供了新的思路。

#### 参考文献:

- [1] Jay Lee, Hung-An Kao, Shanhu Yang, Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment, *Procedia CIRP*, Volume 16, 2014, Pages 3–8, ISSN 2212–8271.
- [2] 彭宇,刘大同,彭喜元. 故障预测与健康管理技术综述[J]. 电子测量与仪器学报,2010,01:1–9.
- [3] Frank P M. Fault diagnosis in dynamics systems using analytical and knowledge-based redundancy: a survey and some new results. *Automatica*,1990,26(3): 459–474.
- [4] Ahmad Alzghoul , Magnus Löfstrand .Increasing availability

of industrial systems through data stream mining[J]. *Computers & Industrial Engineering*,2011.

- [5] 孙冬. 基于混合模型的故障检测与诊断方法的研究与应用[D].南京航空航天大学,2013.
- [6] 刘强,柴天佑,秦泗钊等. 基于数据和知识的工业过程监视及故障诊断综述 [J]. 控制与决策,2010,第25卷(06): 801–807+813.

#### 作者简介:

孔宪光 (1974–), 男, 副教授, 博士。主要研究方向为工业大数据与工业互联网。E-mail: kongxg@vip.sina.com

钟福磊 (1990–), 男, 研究生。主要研究方向为工业大数据环境下的故障预测和健康管理。E-mail: zhongfulei@126.com