（封面、封底用120克白色铜版纸打印，红色字打印时删掉）



**本科生毕业设计[论文]**

（上一行为华文中宋小初号加粗居中）

**基于工业大数据的生产设备故障诊断**

（黑体2号加粗居中）

|  |  |
| --- | --- |
| 院 系 | 机械科学与工程学院 |
| 专业班级 | 机械1401班 |
| 姓 名 | 张照博 |
| 学 号 | U201410606 |
| 指导教师 | 金海、吴波 |

年 月 日（华文中宋3号居中）

**学位论文原创性声明**

（黑体小2号加粗居中）

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包括任何其他个人或集体已经发表或撰写的成果作品。本人完全意识到本声明的法律后果由本人承担。

（宋体小4号）

作者签名： 年 月 日

**学位论文版权使用授权书**

（黑体小2号加粗居中）

本学位论文作者完全了解学校有关保障、使用学位论文的规定，同意学校保留并向有关学位论文管理部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权省级优秀学士论文评选机构将本学位论文的全部或部分内容编入有关数据进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于 1、保密囗，在 年解密后适用本授权书

2、不保密囗 。

（请在以上相应方框内打“√”）

（宋体小4号）

作者签名： 年 月 日

导师签名： 年 月 日

(注：此页内容装订在论文扉页)

**摘 要** （黑体小2号加粗居中）

由于计算机硬件以符合摩尔定律的速度迅猛发展，计算机数据存储、数据传输和分布式计算的成本都大幅度降低。而现代化的工厂中，往往都布置有大量的传感器，而且随着存储成本的降低，读取到的设备信息变得更加丰富，由此便会产生海量的工业数据。

对于这一变化，全世界许多国家都相继提出相应的举措。最初是德国提出了“工业4.0”的概念，之后美国推出“工业互联网”，我国也相继推出“中国制造2025”的概念，其核心都指向智能制造。而工业大数据技术则是这些内容中的核心部分。

在工业设备的运行过程中，自然磨损、设备超载、操作不当等多种原因会导致设备的性能发生下降，甚至于产生故障或者是异常。而通过对设备加装传感器进行监控，获取到设备的实时信息并且加以梳理计算，就可以得到设备各个部位的实时运行状态，从而实现对设备的监控。而如果出现了设备故障现象，则可以通过对历史数据进行数据挖掘、清洗形成故障模型，导入设备的最新运行数据进行故障诊断。

设备的故障诊断方法可以分为三种:基于机理模型的方法，基于数据驱动的方法，基于知识工程的方法[1]。本文将采用基于数据驱动的方法中的基于分类的方法进行故障模型的构建。同时为了对比不同分类方法的性能，本文采用了两种分类方式进行比较。

本文主要研究工作和成果如下：

（1） 建立了以决策树算法为基础的故障诊断模型；

（2） 实现了以支持向量机为核心的数据驱动方法。对支持向量机的核心原理进行了研究，以风力涡轮机齿轮箱的健康数据、故障数据进行了SVM的分类检测，实现了简单的故障诊断；

（3） 比对了两种策略的运行时间，故障诊断正确率以及其他的一些的性能度量。

（宋体小4号，行间距固定1.5倍行距，字符间距为标准）

**关键词：**故障诊断；工业大数据；数据驱动；决策树；支持向量机

（黑体4号加粗） （宋体小4号）

**Abstract**（Time New Roman小2号加粗居中）

（Time New Roman小4号，行间距固定1.5倍行距，字符间距为标准）

As the computer hardware rapidly develops at a rate consistent with Moore's Law, the cost of computer data storage, data transmission, and distributed computing has been greatly reduced. In the modern factories, a large number of sensors are often arranged, and as the cost of storage decreases, the read device information becomes more abundant, and thus a large amount of industrial data will be generated.

For this change, many countries around the world have put forward corresponding measures in succession. Initially, Germany proposed the concept of "Industry 4.0". After the United States introduced the "Industrial Internet," China has also successively introduced the concept of "Made in China 2025." Its core points to intelligent manufacturing. Industrial big data technology is a core part of these contents.

During the operation of industrial equipment, natural wear, equipment overload, improper operation, and other reasons can cause the performance of the equipment to drop, and even result in failure or abnormality. By adding sensors to the equipment for monitoring, obtaining real-time information from the equipment and combing calculations, the real-time running status of various parts of the equipment can be obtained, thereby realizing the monitoring of the equipment. If there is a device failure, you can perform data mining and cleaning on the historical data to form a fault model and import the latest operating data of the device for fault diagnosis.

Currently, there are three types of device diagnostic methods that are popular in the world: mechanistic model-based methods, Data-Driven methods, and knowledge-based methods. This article will use a classification-based approach based on a Data-Driven approach to build a fault model. At the same time, in order to compare the performance of different classification methods, this paper uses two classification methods to compare.

The main research work and achievements of this article are as follows:

(1) Establish a fault diagnosis model based on decision tree algorithm, and study the differences between ID3 and C4.5;

(2) A Data-Driven approach based on Support Vector Machines(SVM) is implemented. The core principle of SVM is studied. The classification and detection of SVM is performed based on the health data and fault data of the wind turbine gearbox, and a simple fault diagnosis is realized.

(3) Compare the running time of two strategies, the accuracy rate of fault diagnosis, and some other performance metrics.

**Key Words：**Fault Diagnosis; Industrial Big Data; Data-Driven Method; Decision Tree; Support Vector Machine

（Time New Roman 4号加粗） （Time New Roman小4号）

**目 录**（黑体小2号加粗居中）

**摘要** Ⅰ

**Abstract** Ⅱ

**1 绪论** 1

1.1 ×××××× 1

1.2 ×××××× 3

1.2.1 ×××××× 7

1.3 ×××××× 10

**……**

**……**

**……**

3**□□□××××××** 20

3.1□□×××××× 20

3.2□□×××××× 23

3.2.1□□×××××× 25

3.3□□×××××× 30

**……**

**……**

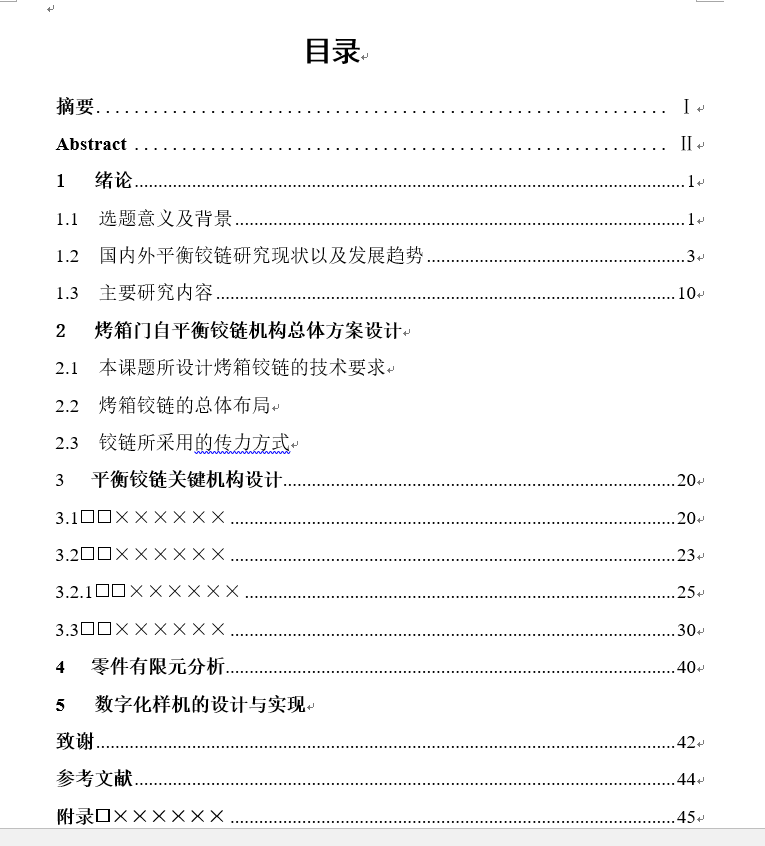
**4□□□结论** 40

**致谢** 42

**参考文献** 44

**附录□××××××** 45

（章为宋体小4号加粗，其余宋体小4号，字母、阿拉伯数字为Time New Roman小4号）

****

**1 绪论**（黑体小2加粗居中）

**1.1 选题背景和意义**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

如今的大数据的数量级，一般都是说是在“太字节”即2的40次方以上的，一般情况下难以收集、存储、管理以及分析的数据。而且随着科技进步，大数据对于“大”的认定还会不断变化。但是大数据不仅仅在于数据量的大小。在计算机行业还未发展到如今这般规模的时候，人们只能选择抽样的数据、局部的数据和片面的数据，纯粹靠经验、理论、假设和价值观去发现未知领域的规律。而这样做的结果，就是对真实现象的抽象归纳与演绎推理，这就不可避免的包含了各种主观上的因素。同时由于样本的局部性，很多推力归纳出来的结果与实际现象具有极大的偏差。

大数据真正的意义在于：我们可以通过传感器，实现与真实世界更加紧密的连接。在得到实时数据后进行整合、挖掘、云计算，去逐步的逼近真实世界，挖掘出那些未曾被我们发现的隐藏规律，建立更加符合真实的数学模型。工业大数据是智能制造的关键技术，主要作用是打通物理世界和信息世界，推动生产型制造向服务型制造转型[2]。

而在工业领域，工业大数据的一大发展方向就是故障诊断，数据的产生和记录贯穿于一台设备从投入生产到损耗的全过程，而通过一定数量的智能传感器，我们可以监控某些生产设备的所有信息，使设备在生产线的实时状态远程监控成为可能，这一方面改善了工作人员的工作环境，另一个更为重要的方面就是提高了设备发生故障或者异常时候的反应速度以及排除故障的效率。

伴随着工业生产水平发展的突飞猛进，工业设备精度越来越高，结构越来越复杂，所以在车间内很多设备的故障都未能得到及时的发现和解决，这一点很容易对工厂造成巨大的损失.由于设备愈加复杂所导致的设备故障信息数据呈现指数型增长，而运行时所产生的海量数据，采用传统基于机理模型的方式已经很难负载如此巨量规模的数据分析，进行故障诊断了。

此外，工业设备结构极其复杂，不同模块之间可能会产生故障的交集，人工分析或者是传统的先验知识故障检测手段已经无法准确、迅速的完成故障的诊断。因此，结合工业大数据对工业设备所产生的海量数据进行数据挖掘分析建立故障诊断模型，对于提高设备维护效率、迅速有效解决故障、降低维修费用有巨大意义。

（宋体小4号，行间距固定1.5倍行距，字符间距为标准）

**1.2 国内外研究现况及发展趋势**

**1.2.1 国内研究现状**（黑体小4号加粗, 字母、阿拉伯数字为Time New Roman小4号加粗）

金风科技股份有限公司基于大量风场的历史故障信息进行大数据分析， 对 SCADA(Supervisory Control and Data Acquisition，监控和数据采集系统)瞬时数据的时间序列模式提取，挖掘桨距角一致性、变桨过程曲线模态、振动模式、变桨电机温差、ng5充电电流差异等断裂征兆模式，通过多模型融合和深度学习，提前90h进行断裂预警，通过预防性维修消除重大故障隐患 [3]。张鹏在在线性模型的基础上进行了大量改进，提出了将卡尔曼滤波器和基于非线性模型相结合的方法，并在航空发动机与传感器上进行了验证[4]。徐德民等人以航行器为研究对象，使用连续-离散无迹卡尔曼滤波算法对航行器执行器进行故障诊断[5]。南京航空航天大学鲁峰等人对发动机进行了仿真建模并且获得了影响系数矩阵，成功实现了对发动机中的气路故障进行诊断[6]。

**1.2.2 国外研究现状**（黑体小4号加粗, 字母、阿拉伯数字为Time New Roman小4号加粗）

国际权威专家 Frank 将故障诊断的方法总结为三种:基于机理模型的方法、基于数据驱动的方法、基于知识工程的方法[7] 。葡萄牙科英布拉大学的Marco S. Reis教授和Geert Gins在《Industrial Process Monitoring in the Big Data/Industry 4.0 Era: From Detection, to Diagnosis, to Prognosis》一文中提到，过去的重点都是检测，也就是基于机理模型的一种对比当前数据的检测办法，实现高水平的检测速度和强度是过去IPM研究的主要重点。在处理新流程时，这是一个必要的步骤，但是随着时间推进，越来越多的因素阻碍了更先进的监控手段的发展[8]。因此为了解决这些障碍与挑战，我们需要找到一种方法可以找到故障根源，也就是我们当今时代的主流过程监测手段：故障诊断。未来的发展方向更是令人心驰神往，故障预检测，能够根据当前的运行数据获取未来一段时间内的设备运行状态预测。

瑞典吕勒奥理工大学的Lianwei Zhang开发除了一套专用于大数据检测以及维护的系统以及一种基于自适应核密度的异常检测（Adaptive-KD）方法，在工业场景中具有极大的使用价值。

**1.3 主要研究内容**

1.3.1研究数据挖掘算法，采集数据进行模型构建，并且根据新的数据进行模型改进、重构；

1.3.2 研究关联度计算算法，对故障数据进行适当处理，提高最后得到的故障模型的精度；

1.3.3 研究机器设备的属性之间的联系，有效的剔除一些对模型精度无益的内容，提高运行效率；

1.3.4 研究模型的改进方案，如决策树中的“剪枝”方法，对模型进行精简，减少模型构建所消耗的资源。

**2 故障诊断的总体设计方案**

（黑体小2加粗居中, 字母、阿拉伯数字为Time New Roman小2号加粗）

**2.1 故障模型的要求**（黑体4号加粗, 字母、阿拉伯数字为Time New Roman4号加粗）

故障模型应该是基于历史数据构建的，由于我们的模型是基于数据驱动，所以对于机械设备方面的先验知识需求量远小于基于机理模型和基于知识工程的构建方式。另外，我们的模型要能接受新的运行数据，对于数据进行测试，从而进行故障诊断的最终目的。

在精度上，模型的精度应该随着数据的不断完善而改进，构建模型所使用的数据越多，那么我们的模型的泛化程度就越高，对于各种实际情况的解读能力就会进一步提升。

**2.2 决策树建立故障树模型**

**2.2.1 信息熵**

假设当前样本集D中有N个样本，而整个样本有k个分类，每个分类对应的样本数量为那么对于每个分类，他们各自占据的信息量（也可以理解为样本分类的频率）为：

结合每一个分类的信息量，则此样本总体的信息熵为：

**2.2.2 信息增益**

假设当前样本集D中有N 个样本，每个样本都有一些属性，假设我们目前取属性A作为我们计算信息增益的属性。

根据属性A，我们可以属性A的不同取值（假设有v种），将整个样本集D分为v个子样本集，每个样本子集的样本数为N，那么每一个样本子集的频率为：

那么该样本集的A属性的信息增益即为：

下面是利用这两个概念获得故障树的过程：



图2 ID3算法生成决策树流程图

**2.2.3 ID3算法**

×××××，其×××××可表示如下：

 (3-1）

 (3-2）

×××××××××××××××××××××××××××× (如表3-1所示)

**2.3 支持向量机二分类模型**

**2.3.1 SVM原理**

基于支持向量机（SVM）的故障诊断方法将诊断问题看成样本分类的问题，即根据历史数据训练出分类器。将数据空间划分成不同的区域，每个区域对应一种运行状态，然后将测试数据投影至数据空间。通过定位其所在区域，推测出测试数据对应的运行状态。下图是一个三分类样例。

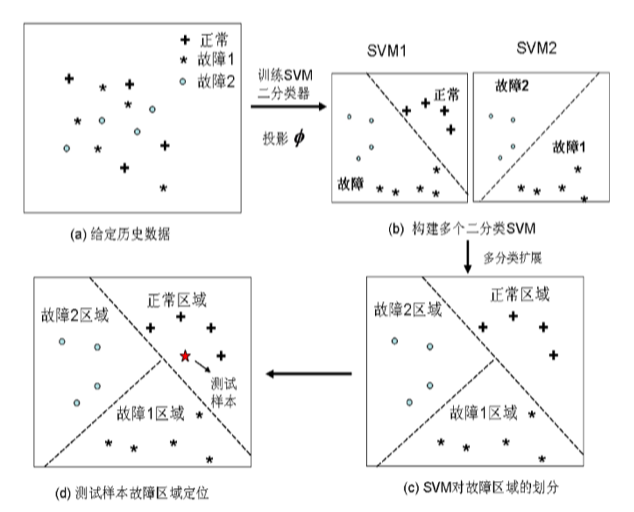


图3 三分类分类样例

在图中，针对目标系统，该方法首先采集各种运行状况下的数据。构建出训练样本集: 然后对训练样本集进行数据预处理，包括去量纲化、特征选择等。并采用SVM对处理过的数据进行学习，生成整个二分类器。见图(b); 由于故障诊断、尤其是故障隔离通常需要面对多分类问题，所以还需要用特定的多分类扩展策略。将多个: 二分类器组合成一个整分类器，使得SVM能够区分多个故障区域，见图(c) 最后，当故障区域被有效划分时。我们只需要对测试样本进行投影确定该样本所属区域，即可实现故障隔离。

SVM的成功运用到实际主要得益于两大技术: 其一，依据SRM准则设计间距最大的分类超平面: 在高维空间里计算出线性最佳分类面。其二，根据核函数媒介得出输入空间里非线性学习算法。关于核函数的方法是当前比较活跃的研究领域。核函数的方式就是通过非线性变换将非线性空间里数据样本集映射至高维线性空间里，在高维空间里导求线性学习算法，要是各个坐标分量之间的相互影响仅仅局限于内积时，那就不用知晓具体非线性变化地形式，只需要把满足Merce条件的核函数代替线性算法里的内积，即可得出原先输入空间里的非线性算法心调。

常用到符合Mercer规定地核函数包括多项式函数、径向基函数以及Sigmoid函数。选取一个核函数就会得出一个SVM。

**2.3.2 SVM核函数**

**3 具体方案设计**

×××××××××××××××××××××××××××××××××××

表3-1□××××××××××

|  |  |  |  |
| --- | --- | --- | --- |
| ××××× | ××× | ××× | ××× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |
| ××××× | ×× | ×× | ×× |

（表标题：位于表格上方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号，表内容：宋体5号，字母、阿拉伯数字为Time New Roman 5号）

××××××××××××××××××××××××××× (如图3-1所示)



图3-1□××××××××××

（图标题：位于图下方，黑体小4号，字母、阿拉伯数字为Time New Roman小4号）

.......

--------章与章之间插入分页符----------

**参考文献** (黑体小2号加粗居中)

[1] 钟福磊. 工业大数据环境下的混合故障诊断模型研究[D].西安电子科技大学,2015.

[2] 卫凤林,董建,张群.《工业大数据白皮书(2017版)》解读[J].信息技术与标准化,2017(04):13-17.

( 宋体小4号)

[3] 王建民.工业大数据技术[J].电信网技术,2016(08):1-5.

[4] 张鹏. 基于卡尔曼滤波的航空发动机故障诊断技术研究[D]. 南京: 南京航空航天大学, 2009.

[5] 徐德民, 刘富樯, 张立川等. 基于改进连续-离散无迹卡尔曼滤波的水下航行器故障诊断 [J]. 西北工业大学学报, 2014, 32(5): 756-760.

[6] 鲁峰, 黄金泉, 孔祥天. 基于变权重最小二乘法的发动机气路故障诊断[J]. 航空动力学 报, 2011, 26(10): 2376-2381.

[7] Frank P M. Fault diagnosis in dynamics systems using ana-lytical and knowledge-based redundancy: a survey and some new results[J]. Automatica, 1990, 26(3): 459-474.

[8] Reis M S, Gins G. Industrial Process Monitoring in the Big Data/Industry 4.0 Era: From Detection, to Diagnosis, to Prognosis[J]. Processes, 2017, 5(3): 35.

……

……

**附录** (黑体小2号加粗居中)

×××××××××××××××××××××××××××

( 宋体小4号)

……

……

……