工业大数据环境下的混合故障诊断模型研究

作者姓名 钟福磊

指导教师姓名、职称 保宏 教授

申请学位类别 工学硕士

学校代码 10701　　学　号 1304122003
分 类 号 TH17　　密　级 公开

西安电子科技大学

硕士学位论文

工业大数据环境下的混合故障诊断模型研究

作者姓名：钟福磊

一级学科：机械工程

二级学科：电子机械科学与技术

学位类别：工学硕士

指导教师姓名、职称：保宏教授

学　　院：机电工程学院

提交日期：2015 年 12 月

A study of hybrid modeling technique for fault detection based industrial big data

A thesissubmitted to

XIDIAN UNIVERSITY

in partial fulfillment of the requirements

for the degreeof Master

in Electromechanical Science and Technology

By

Zhong Fulei

Supervisor: Bao HongProfessor

December 2015

西安电子科技大学
学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名：_____　　日　　期：_____

西安电子科技大学
关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在年解密后适用本授权书。

本人签名：_____　　导师签名：_____

日　　期：_____　　日　　期：_____

摘要

近年来，随着互联网、物联网、传感器等信息技术与通信技术的迅猛发展，数据量的暴涨成了许多行业共同面对的严峻挑战和宝贵机遇。随着工业中信息技术的进步和现代化管理理念的普及，企业的运营越来越依赖信息技术。在现在的工业中，已经存储了大量的设备工况数据，已经呈现出大数据的诸多特征，但是企业并没有挖掘出这些数据的价值。现在信息通信技术不断融入工业设备中，推动工业设备向数字化、智能化方向发展。具体而言，在设备、生产线中配备传感器，抓取数据，然后通过无线通信连接互联网，传输数据，对设备或生产过程进行实时状态监控。

在设备运行过程中，自然磨损或者意外事件会使设备的性能发生一定的变化，会出现故障。现在可通过传感技术感知数据，通过对设备运行过程中的各个因素精确感知并进行分析来实现设备的故障诊断。在一定程度上，设备所产生的数据直接决定了设备智能化水平和故障诊断的准确性。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。为了充分利用工业中海量数据并且保证诊断的准确性和高效性，本文将基于数据驱动的方法和基于知识工程的方法结合起来，提出了一种混合故障诊断模型。盾构机故障诊断中该方向研究较少，所以本文选择盾构机为研究对象，对其管片拼装机的液压回路进行仿真，利用仿真数据对混合模型中的关键算法进行研究。

本文具体工作如下：

（1）提出了基于数据驱动的方法和基于知识工程的方法相结合的混合故障诊断模型。

（2）提出了以支持向量机算法为核心的数据驱动方法。对支持向量机算法涉及到的核函数和多分类等问题进行了研究。以盾构机中拼装机的液压系统为研究对象，在 AMESim 软件进行仿真，研究油液污染和漏油两种故障情况，通过对马达的运行数据进行分析来进行故障诊断。

（3）将数据挖掘算法应用到基于知识工程的方法中进行知识发现。本文分别应用关联规则算法和聚类算法于盾构机故障仿真数据进行知识发现。关联规则算法采用了 Apriori 算法，聚类算法采用了 k-means 算法。

关 键 词：工业大数据， 基于数据驱动的方法， 基于知识工程的方法， 故障诊断，盾构机

ABSTRACT

In recent years,with the rapid development of the Internet, Internet of things, sensors and other IT and communication technology. Many factories have to face the challenges of the data volume soared. With the popularity of technology and modern management concepts in industry, companies operations become increasingly dependent on information technology. Industry has stored a large number of equipment condition data and has shown a lot of characteristics of big data, but companies do not find out the value of these data. Specifically, it can put sensors into the equipment and production line, and then connect to the Internet to transmit data via wireless communication. In this way, it can realize production of equipment for real-time monitoring.

During operation period of devices, natural wear and tear or accident will make the performance of devices changed. Now use the available of sensor technology and real-time sensing data, the running data of the devices will be perceived and realize fault diagnosis.
The fault diagnosis method can be divided into three types: data-drived method, first-principle based methods, knowledge-based methods. Based on the industry big dataand meet the requirements of reliability, this paper combine the knowledge-based methods and data-driven methods present a hybrid model of fault diagnosis and used the hybrid model in shield machine hydraulicsystem fault diagnosis.

In this paper, the works are summarized as follows:
(1)Firstly, this paper proposed a hybrid model for fault diagnosis.
(2)Secondly, this paper proposed the data-drived method based on support vector machine algorithm. We analysisthe kernel function and multi- classification problems. This paper used this method in hydraulic systemof shied machine and in AMESim simulation software to research oil pollution and oil spill fault.
(3)Finally, this paper apply data mining algorithms to discover the knowledge. In this paper, we use Apriori algorithm to extraction rules in mass data and apply cluster algorithm k-means to find rules.

Keywords: Industrial Big Data, Knowledge-based Method, Data-drived Method, Fault Diagnosis, Shield Machine

插图索引

表格索引

符号对照表

符号	符号名称
P	规则产生中的条件
Q	规则产生中的结果
R_{emp}	经验风险最小化
f	预测函数
L	损失集
(x_i, y_i)	训练样本
N	样本数
ω	广义参数
$\boldsymbol{w}$	支持向量机算法中参数
b	支持向量机算法中参数
α	Lagrange 中对偶参数
$\boldsymbol{w}^*$	$\boldsymbol{w}$ 的最优值
b^*	b 的最优值
$\boldsymbol{a}^*$	经 Lagrange 变换后支持向量机的最优解
$K(x_i, x_j)$	核函数
Φ	非线性函数
ξ	松弛因子
σ	尺度参数
$f_k(x)$	选择函数
$\mathrm{sgn}[f_k(x_i)]$	$f_k(x_i)$ 所对应的的类别
m	选择函数最大值所对应的类
$k_{\max}$	最大频繁项集长度
$s(X \to Y)$	支持度
$c(X \to Y)$	置信度
μ_j	类的质心
$c^{(i)}$	k-means 算法中样本所属的类
$\min \left\| x_j^{(i)} - \mu_j \right\|^2$	样本到质心的最小距离
$J(\mathrm{c}, \mu)$	畸变函数

缩略语对照表

缩略语	英文全称	中文对照
OCSVM	One Class Support Vector Machine	单一类支持向量机
PCA	Principal Component Analysis	主成分分析
SVM	Support Vector Machine	支持向量机
ERM	Empirical Risk Minization	经验风险最小化
SRM	Structural Risk Minimization	结构风险最小化
RBF	Radial Basis Function	高斯核函数
DM	Data Mining	数据挖掘

目录

第一章　绪论

1.1 论文研究背景

本论文以工业大数据为背景，研究复杂设备的故障诊断。由于互联网的迅速发展，数据传输、数据共享和分布式计算的成本都大幅度降低。在现代化的工厂中传感器布点越来越多，数据获取也变得十分方便，在大型设备中收集到的工况数据在不断的增加，于是产生了海量的工业数据。针对工业中的这种变化，德国政府提出了一个科技战略计划，称之为工业 4.0，并认为工业 4.0 将会带动第四次工业革命，是继蒸汽机、电气、信息化后的又一次工业浪潮。

大数据是工业 4.0 时代的重要特征，大数据在工业界兴起的主要原因有：（1）获取实时数据的成本已经不再高昂；（2）设备自动化过程中产生了大量的数据，这些数据所蕴藏的信息和价值并没有充分挖掘；（3）嵌入式系统、低耗能半导体、处理器、云计算等技术的兴起使得设备的运算能力大幅提升，使计算机具备了处理实时大数据的能力；（4）制造流程和商业活动变得越来越复杂，依靠人的经验和分析已经无法满足如此复杂的管理和协同优化的需求。美国政府也对工业 4.0 比较重视，他们认为其价值主要有：（1）利用大数据分析技术发现隐藏的信息或风险，使以往不可见的风险能够被避免；（2）实现产品的智能化升级，利用实时状态监控系统，对设备实现智能诊断和预防维修。

百度、谷歌等互联网巨头在一天内会产生大量的数据，但据美国一家公司对美国各个行业的数据量进行了估计后发现位居首位的是离散制造业。举一个例子，飞机汽轮压缩器叶片一天就会产生 588GB 的数据，而世界上最大的微博公司(Twins)每天才产生 80G 的数据，由此可见飞机上区区一个汽轮压缩器的数据量是一个互联网巨头的 7 倍之多。所以工业中的数据分析就显的十分的重要，而且具有广阔的天地。现如今企业并没有很好的利用这些数据，用这些数据来对机械设备进行故障原因、严重程度等的判断，这些数据的利用率是非常低的。Jay Lee 教授认为全球化是本地工业面对的最大挑战，同时也是制造业必须面对的问题，所以要保证制造业在全球具有强大的竞争力，必须将制造业推向下一个转型。为了保证制造业的竞争力，制造业必须引进更多先进的技术，如先进的分析技术、物理信息系统等来改善制造业的效率和生产能力。将整个制造业向物联网方向转变，数据将会变得非常重要，这就需要开发出性能优越的数据分析工具。Jay Lee 教授对工业大数据也有其独到的见解，他认为：“工业大数据就是将一系列的机器连接在一起，应用先进的分析工具就能系统的将数据转换为信息，这样在处理一些不确定事件时候，就可以做出正确的判断”[1]。Jay Lee

教授希望机器设备能够具有自我意识（self-awareness）、自我预测(self-prediction)、自我比较(self-comparison)、自我识别(self-reconfiguration)和自我维护(self-maintenance)。

在实际应用过程中，面对海量数据主要利用数据挖掘和机器学习等智能算法来进行关联规则、分类、聚类等分析。这些算法对数据的分析能力远远超过了人类，而且克服了传统的知识发现方法的缺点，形成了初步的全自动式知识发现。例如，中国台湾高盛的带锯机床就是工业大数据进行价值创造的典型实践案例。高盛是生产带锯机床的企业，所生产的带锯机床产品主要用于对金属物料的粗加工切削，为接下来的精加工做准备。机床的核心部件就是用来切削的带锯，带锯的磨损严重影响切削质量。高盛利用 PLC 控制器和外部传感器收集的数据，收集了大量的带锯全生命信息档案并形成了一个庞大的数据库。王国彪等人针对机械故障诊断的发展方向进行了认真的思考，他们认为故障诊断技术虽然发展的很迅速，但是仍然有很多的不足，认为在理论、方法、智能化等几个方面一直没有突破性进展。他指出现在智能诊断系统还很薄弱，随着机械设备的大型化、复杂化、高速化、自动化和智能化，迫切需求融合智能传感网络、智能诊断算法和智能决策预示的智能诊断系统、专家会诊平台和远程诊断技术等[2]。

现如今城市中的人口越来越多，尤其是一线城市，所以必须快速发展城市交通，故各大城市都在大力修建地铁。修建地铁时候，主要依靠盾构机设备来进行掘进，盾构机是大型复杂设备，国内一般都依靠进口德国和日本的产品。盾构机主要应用于地铁隧道、铁路隧道、公路隧道等各种隧道工程。盾构机结构十分复杂，而且工作环境恶劣，较容易发生故障，如果能够及时发现盾构机的故障甚至能够预测到整个设备的未来运行状态就十分的关键。若能在盾构机中加入合适的传感器，这样在盾构机运行过程中，就能够通过传感器来收集到大量的工况数据，然后再通过专业的软件或专家人员对这些数据进行分析，能够实现对盾构机的故障诊断，甚至能够预知盾构机的故障，降低其运营成本。

1.2 故障诊断研究现状

在新闻中会经常看到一些因为设备故障而造成的惨案，轻则造成一定的经济损失，重则造成重大人员伤亡，所以故障诊断的研究具有十分重要的意义。2000 年三峡“九.三”事件，由于塔带机断裂而造成了三十三人的伤亡事件。在我们的日常生活中屡见不鲜的电梯事故，即使是在科技实力领先的航空航天设备中，故障也是难免的，如航天设备“玉兔”就曾经出现过故障。如果故障诊断技术能够在故障的萌芽阶段就将其判断出来，会减少很多重大的灾难性事故。

国内外有很多的专家研究故障诊断，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[3]。基于机理模型的方法提出的比较早，主要是建立一个精确的机理模型，然后利用数学方法来对输出数据进行分析。基于知识的方法主要是根据历史先验知识，按照相应的算法来对故障现象或者故障数据在知识库中进行搜索匹配，寻找出故障。基于数据驱动的方法是近几年兴起的一种方法，工业中能够的采集到大量的系统运行过程数据，这些数据都与设备的工作状态的相关，这些数据是工业财富。该方法通过利用采集到的输入输出数据，分析数据的各种统计特征，建立过程的数据特征模型。

基于机理模型的方法研究的时间较早，案例也很多。张鹏等人在对线性模型的研究基础之上进行了改进，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，应用连续-离散无迹卡尔曼滤波算法来对航行器的执行器进行故障诊断[5]。南京航空航天大学鲁峰等人对某发动机型号进行了仿真建模并且获得了影响系数矩阵，对发动机中气路故障进行诊断[6]。

基于知识工程的方法基本上分为图论法和专家系统两种方法。华南理工大学刘其洪等在INV1612试验台对转子进行了研究，开发了专门针对该设备的专家系统[7]。陈超等人针对数据库中利用多源信息的不足的问题，将工艺信息加入其中。该方法加强了工艺信息对机械故障诊断的影响，也说明了知识库中存储的设备相关信息越全面，诊断准确性越高[8]。盛博等人建立了数控机床多故障模型，利用图论方法进行诊断[9]。

国内一些专家已经在如何利用数据驱动方法进行故障的诊断有了一定的研究成果。清华大学的萧德云等人将数据驱动方法具体进行了分类，他们还对数据驱动方法的背景以及和其他方法之间的区别进行了研究，通过其论文可以了解到各种故障诊断方法[10]。Xuewu Da等人介绍了复杂系统中利用数据处理来进行故障检测和诊断。根据数据的类型以及对数据的加工方式，将故障检测诊断系统分为三类：基于历史的数据驱动方法、在线数据驱动方法和信号法，并对未来工业中自动化的来检测故障进行了畅想[11]。马贺贺等针对实际工业过程中存在的诸如非线性问题、多模态问题、过程数据复杂分布的问题展开研究工作[12]。瑞典吕勒奥大学Ahmad Alzghoul等人对如何应用数据挖掘技术来提高工业设备的可靠性进行了研究，主要对单一类支持向量机(OCSVM)、基于多边形(polygon-based method)方法和基于网格的方法(grid-based method)进行了比较。论文以Bosh Rexroth Mellansel AB型液压设备为研究对象，能够在设备故障发生早期就能够检测出来，并进行诊断[13]。爱荷华大学智能系统实验室Andrew Kusiak等人主要将数据驱动方法应用在风轮机的故障诊断和预测中，应用了五种方法来建立故障处理模型。他们对风轮机中叶片变桨进行状态监控，对变桨的两个故障进行分析，通过数据来发现这两个故障之间的关系以及它们对整个风轮机的影响[14]。辛辛那提大学Jay Lee等人早在2007年左右就开发出了一套智能预测性诊断和

维修工具。智能维修主要是要在故障发生之前就将发现故障，将其解决掉，减少工厂无计划停机时间[15]。

1.3 盾构机故障诊断研究现状

盾构施工效率十分高、一次就能成洞，广泛应用与城市中隧道修建中，其构造复杂，涉及到的组件设备较多，较易出现故障。例如：2015年2月11日武汉发生地陷，道路发生严重塌陷，就是由于盾构机尾部发生故障，而工作人员恰好外出。2009年南京长江隧道工程盾构机在进行施工作业的时候，一个刀盘辅臂舱内突然发生了泥浆泄露，盾构机启动自保护模式，两名工作人员被困。每年在我国以及世界各地都会有大量的盾构机发生故障，做到及时诊断，甚至能够预测出其可能发生的故障具有十分重要的意义。

盾构机在施工过程中，当地的地质和周边环境对其工作状态、性能评估以及故障诊断等都有着严重的影响，所以国外的研究在我国的施工过程中很难进行应用，我们必须根据施工的具体环境来进行判断。如在俄罗斯冰冻地质环境下进行盾构施工与在我国的平原地区施工，整个地质环境、气候环境等差别都是比较大的。而且俄罗斯的盾构机和我国的盾构机型号也是有区别的，他们的更加庞大。石家庄铁道大学左庆林等分别应用振动、油液、温度等检测方法来对盾构机的核心设备进行故障诊断。他们设计了盾构机故障诊断系统中的软件部分，将仿真数据和实验数据综合起来作为数据集，对系统进行了测试[16]。中铁隧道集团专用设备中心赵华、苏东等人主要对盾构机的主变速箱进行研究，对振动和油质技术进行了研究[17]。韩超等在论文中应用神经网络、最小二乘法等状态预测进行了一定的研究，有助于盾构机的维修[18]。

通过对国内外盾构机的故障诊断文献阅读以及对工业大数据相关知识进行研究后，发现将人工智能技术应用到盾构机中是一个重要的发展方向。盾构机的智能诊断研究在国内外的研究都还是比较的少，即使有些文献中提到了专家系统，但是大多数都是针对某一个特定型号的设备，并不具有通用性。

1.4 混合模型研究现状

在工业中一方面对于复杂设备已经无法凭借机理建模的方法建立一个准确度很高的物理模型，另一方面是现在能够收集到设备运行时候的工况数据，基于数据驱动已经逐步开始成为研究的热点。从2008年开始，IEEE就开始举办The IEEE Int Workshop on Defect and Data Driven Testing，该活动的目的主要是为了解决基于数据驱动的故障诊断技术。

基于机理模型的方法主要是依靠被诊断系统精确的机理模型，将机理模型的输出

与实际输出进行比较，并采用数学方法对残差进行分析处理从而实现故障诊断。它能够反映设备的内部结构和机理，揭示事物内在规律，能够了解故障发生的原因。缺点是当面对比较复杂的设备系统时，无法获取内部机理的全部信息，很难去建立一个精确度很高的模型，该方法过于依赖设备的精确模型[19]。机理分析或多或少都会基于一定的假设和简化，有些设备即使简化之后，仍然较为复杂，而且经过简化后，出现误差会影响判断的准确性。基于知识工程的方法相对比与机理建模方法不需要建立机理模型，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。缺点是通用性差，方案对策不确定，必须通过积累大量的知识来建立“知识库”。当出现一种新的未知故障，由于没有该故障的任何知识，则对故障无法进行判断。基于知识工程的方法优点是它不需要对系统进行数学建模，一般会引入较多信息，如设备定性的结构知识、专家知识、决策知识等。基于数据驱动的方法建立起来更为方便，不需要过程的模型或先验知识，回避了建立对象机理模型的问题，只需要对设备的工况数据进行处理与分析。缺点是由于该方法不考虑系统的机理和内部结构等先验知识，所以建立的数据模型中变量参数不具备明确的意义。三种方法的对比图如表1.1所示：

表1.1 三种故障诊断方法比较

方法	优点	缺点
基于机理模型的方法	能够反映过程汇总的内部结构和机理，揭示事物内在规律，能够了解故障发生的原因。	当产品系统过于复杂，就很难建模。
基于知识工程的方法	只需要具有经验的工程师，将其故障处理经验知识作为先验知识，容易得到，多用在一些模型机理不清楚的情况下。	通用性差，方案对策不确定，难以处理未知情况。
基于数据驱动的方法	方法简单，仅仅需要各种状态下的运行数据。	变量的物理意义不清楚，模型一旦建立，外延性差。

刘强等对基于数据驱动的方法和基于知识工程的方法进行了讨论，对这两种方法的优缺点以及两者是否能够进行结合进行了可行性分析[20]。Ahmad Alzghoul等将数据驱动方法和基于知识工程的方法进行了对比。基于知识工程的方法主要是基于因果模型（故障树），基于数据驱动的方法主要应用了主元素分析法（PCA）。将两种方法应用在单独的工业设备中，对一系列的系统进行监控，这两种方法都能增加工业系统的可用性[13]。Gorka Azkune,Aitor Almeida等为了克服基于知识工程的方法的缺陷，将基于数据驱动的方法应用到基于知识工程的方法中，提出了一个改善模型，他们研究的对象是人类的行为。他们提出的方法包括一种新的聚类处理模型，通过专家知识来初步建立模型，通过这些模型来对行为集进行处理。基于这些行为数据集，设计了

一种新的处理方法来跟踪行为的变化，建立准确的模型[21]。

经过对国内外相关知识的查阅后，大部分研究人员都是应用单一建模方法来对故障进行处理。虽然有一些研究人员在研究混合模型的故障处理方法，但是大多数都是在研究基于机理模型的方法和基于数据驱动的方法相结合的混合建模方法。现阶段将基于数据驱动的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很少有涉及，目前还没有成熟的理论研究，所有本文具有一定的研究价值。

1.5 论文研究的目的和意义

本文主要依托教育部项目基于工业互联网的离散制造业大数据分析与协同管理研究以及科研组与中铁一局的合作项目。研究课题是在工业大数据环境下，针对目前城市轨道交通盾构法隧道施工而研发的信息化系统，为施工过程的风险防范提供可视化的实时工况数据，对盾构机进行智能化监控，为盾构机再制造提供大数据积累与分析。利用该系统解决盾构施工信息共享，控制盾构机施工风险，预知盾构机故障，降低其运营成本。本文的主要研究重点是如何对盾构机进行故障诊断，积累诊断经验，提高设备的诊断效率。该方法主要利用了设备运行过程中的数据，以一个全新的角度来思考故障诊断问题。本文提出的混合模型，不仅仅可以应用在盾构机设备上，还可以推广到其它的机械设备中，是一个通用的模型。将基于数据驱动的方法和基于知识工程的方法进行结合的研究较少，本文具有一定的研究意义。

本文将两种方法结合起来原因有以下两点：首先，在盾构机中该研究才刚刚起步，故障诊断的准确性和效率还有待进一步的提高，整个体系还不成熟；其次，若对每次收集到的数据都应用基于数据驱动的方法来进行判断会影响诊断的时效性，故应用基于知识工程的方法来提高其诊断速度。本文主要研究智能装备的故障诊断，是研究整个工业大数据中十分重要的一部分。整个工业也是由一个个复杂的智能设备组成的，若不对这些设备进行研究，则工业大数据就会成为空谈。

1.6 论文结构组织

本文的创新之处在于将基于知识工程的方法与基于数据驱动的方法相结合来进行故障诊断。在本文中提出了新的知识发现方法，突破了传统知识发现和获取的瓶颈，实现知识的自动获取，实现了知识库的动态更新。本文的研究对象是中铁的大型设备盾构机，在国内针对盾构机进行故障诊断的研究也是比较少的，尤其是以工业大数据的角度来对其进行分析。

论文章节内容安排如下：

第一章介绍了工业大数据的背景，故障诊断的基本方法以及混合模型的思想，并

介绍了故障诊断、混合模型的研究现状。

第二章首先介绍了一种基于工业大数据的混合模型以及建模过程，其次介绍了混合模型中的关键技术。

第三章对支持向量机理论进行了介绍，对其中涉及到的核函数选择以及其参数的确定、多分类等问题进行了研究。

第四章主要介绍了如何利用数据挖掘算法来实现自动发现知识。基于知识工程的方法中选择了专家系统，提出了应用关联规则和聚类算法来自动发现知识。

第五章简单介绍了盾构机的基本原理，并模拟运行时的工况数据。利用基于数据驱动的方法来进行故障诊断以及利用关联规则、聚类等算法进行知识发现，对关键算法进行了研究。

第六章对全文进行总结，指出文中待完善部分。

第二章　混合故障诊断模型

2.1　混合故障诊断模型

混合故障诊断模型是将基于知识工程的方法和基于数据驱动的方法相结合。该模型利用数据挖掘方法进行知识的自动获取，建立知识库。整个工作流程为：首先通过基于知识工程的方法对输入故障数据进行判断，若能利用知识库中的知识对故障进行搜索匹配成功，则整个过程结束，输出结果。若利用基于知识工程的方法判断失败，即知识库中未存储相关故障信息，则利用基于数据驱动的方法来进行判断，得到结果后利用数据挖掘方法进行知识发现。总体框架图2.1所示：

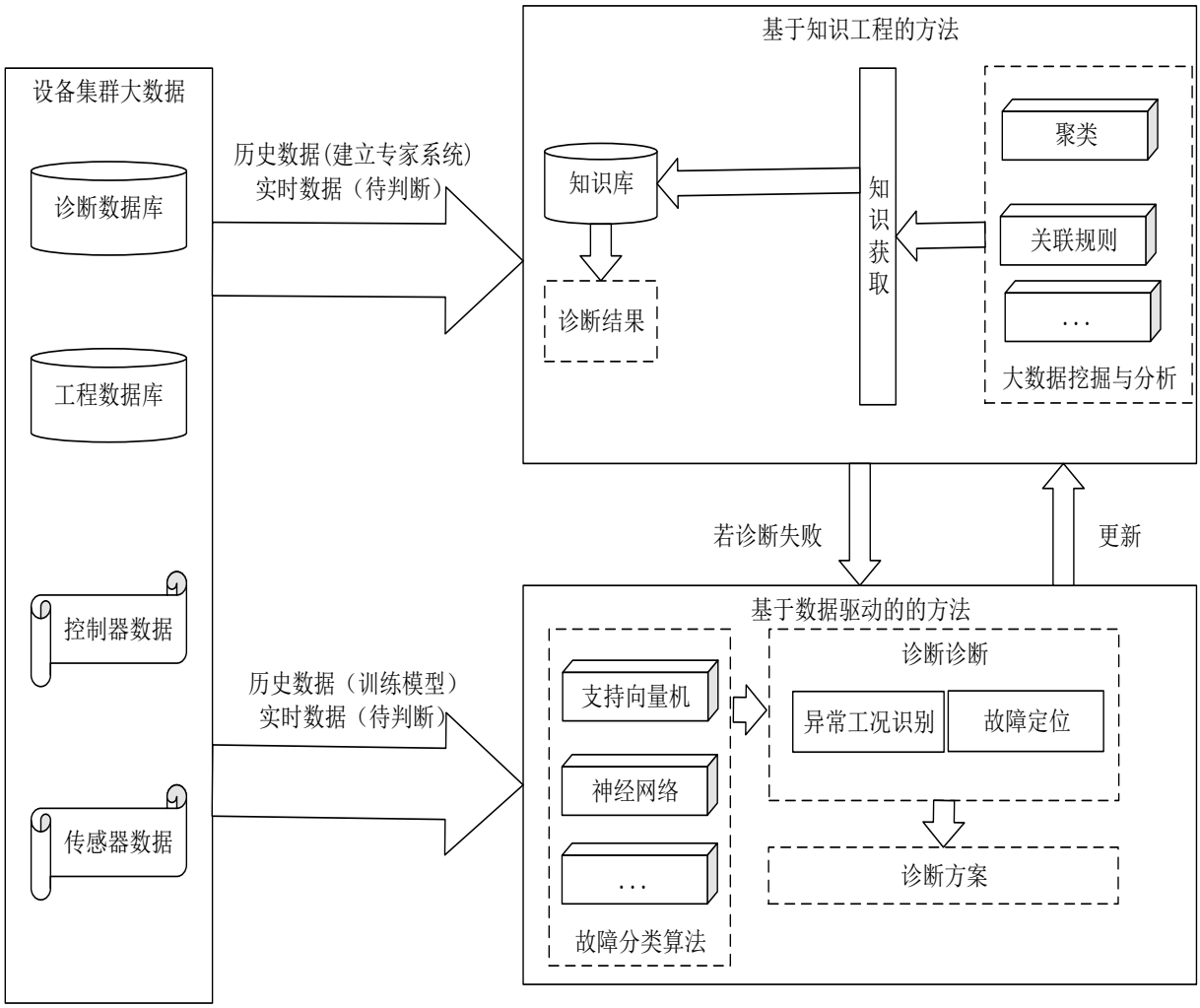

图2.1 混合故障诊断模型图

混合模型的流程如图2.2所示：

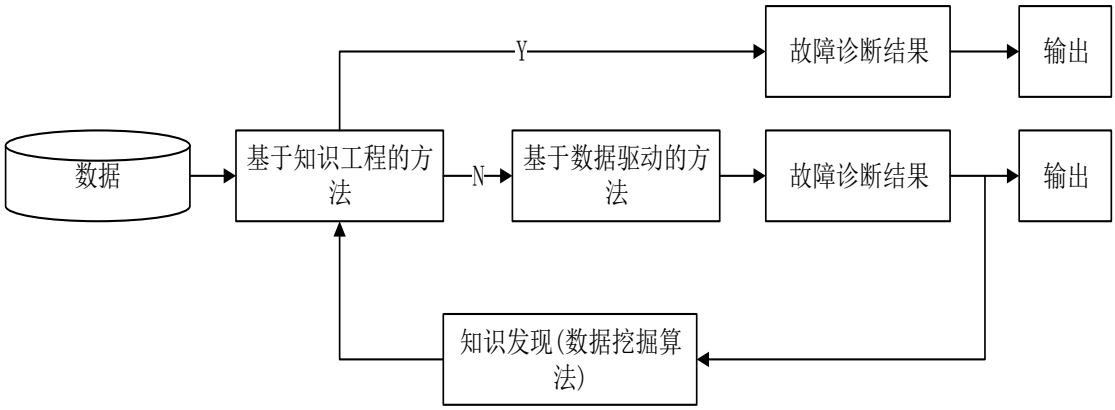

图2.2 故障诊断总体流程图

整个执行过程步骤为：(1)首先尽可能全面的去采集设备状态的相关数据，建立设备数据集。为了得到充分的数据源，需要监测设备工作过程中的状态，采集被监测设备相关的尽可能多的物理量。数据的采集可以通过设备的PLC来进行收集，也可以在感兴趣的部位通过传感器的布点来采集。若想使诊断的准确性很高，可以将设备的设计、建模、工艺、加工、测试、维护数据、产品结构、零部件配备关系等数据也进行收集;(2)利用历史数据建立知识库和分类器。随着设备机械化以及传感器应用程度的提高，设备工作状态信息能够被采集和存储，这为日常数据分析和故障诊断提供了很好的数据源。在这些数据中，首先利用数据挖掘算法进行知识发现，建立知识库。另外建立一个分类器，能够对设备的状态进行判断，在建立分类器时要利用历史数据来对用到的算法进行训练;(3)故障诊断。对于新到达的实时数据，首先利用基于知识工程的方法来对相应的故障进行诊断，若判断成功，则输出故障信息。若判断失败，则利用基于数据驱动的方法来进行判断。最后利用数据挖掘算法来对诊断结果和数据集来进行知识发现，不断丰富知识库。

2.2 混合模型中的关键技术

2.2.1 数据预处理

数据预处理在建立和实现整个模型时具有十分重要的作用和意义，很多的数据挖掘结果不理想都是由于数据的质量不高。在对工业中机械设备进行数据收集时候，由于施工环境的恶劣，收集到的数据肯定包含有大量的错误和噪音。以盾构机中的数据为例子，有的研究人员是通过盾构机 PLC 来得到数据，这些数据与实验室中的数据差别十分大。一个成功的故障诊断模型，不仅仅要保证所选择的算法是当前最合适的，

而且也要保证输入的是能准确反映实际工况的数据。数据预处理的作用就是使输入模型中的数据尽可能的逼近实际设备的运行情况。在对数据预处理之前，样本的确定很重要，合适的样本大小能够控制误差、提高精度。在对大数据的分析中，不可能对所有的样本进行分析，必须要通过一定的方法来选择样本数据。这些样本数据要能够基本代表要处理数据的信息，若样本选择有误则数据挖掘、分析都没有任何意义[22]。原始数据中包含大量的杂音和缺失值，它们并不符合挖掘算法进行知识发现所要求的规范和标准，一般这些原始数据具有以下特征：不完整、含有噪音、杂乱。数据预处理就是在对数据进行分析和挖掘之前对原始数据进行必要的清理、集成、归约和分割等工作。

数据清理：顾名思义数据清理的作用就是清理样本数据中的错误，具体包括对缺失值的补充、对噪音数据的处理、对孤立点的识别和删除、对数据中不一致的问题的解决[23]。在这一步会用到一些算法：如决策树、回归法、贝叶斯等。

数据集成：在对设备进行数据收集的时候，所有的数据不可能来自于一个部件，收集到的一定是数据源较杂的数据，会包含不同的数据格式。这样在综合数据库中，同一个事物或者同一个变量表达的方式就会出现不一样的情况，所以要利用数据集成来将这些杂乱的数据源进行统一化整理，方便进一步的处理[24]。

数据归约：数据归约一般应用特征提取和特征选择，其目的是将一个相对较大的数据集浓缩为一个较小的数据集，但仍然需要保持原来大数据集信息的相对完整。一般的数据集中都会包含大量的属性，经过归约处理后的数据集对分析结果几乎没有影响。两种方法都是为了达到数据降维的目的，但是这两种方法之间也是有一定的差别的。特征提取是将数据集中的属性进行变换得到一系列新的特征，而特征选择是在原始数据集中的属性中进行选择，选择有意义的属性。进行数据归约的时候，应用比较广泛的是主成分分析法(PCA)。PCA 最早是有 Pearson 于 1901 年提出来的[25]，将多个相关变量转变为几个不相关的变量的线性组合，是一种线性映射方法。数据集中的原始属性通过 PCA 算法处理将会出现新的综合变量，新的综合变量能够保证众多变量的信息完整，这些新形成的变量彼此要保证不相关[26]，将包含信息最大的称为第一主成分。如果经过计算判定第一主成分不能够将原始数据集中的大部分信息反映出来，则会在其他的变量中选择一个作为其补充，称为第二主成分。依次轮推，直到这些主成分能够达到要求为止。

数据分割：经过前面几步的处理，已经得到质量较高的数据集了。然后进行数据的分割，数据分割的主要目的是为了保证建立的分类器具有较高的分类准确率，将数据集分为训练集(train set)和测试集(test set)。训练数据集用来对分类器中的参数进行训练，在建立一个分类器时，不可能保证其参数能够达到最优，所以要利用数据来进行训练。测试集是判断训练的结果是否能够达到要求的标准。数据驱动模型中会用到

大数据分析算法，这些算法中的参数都是要通过训练来达到最优的。

2.2.2 基于知识工程的方法

基于知识工程的方法主要是利用与设备故障诊断相关的先验知识，主要包括系统的一些故障集合、故障与现象之间的关系等。利用这些先验知识，建立一个针对所研究设备的定性模型，一旦出现故障，就按照事先定义好的算法来进行搜索，从而得出故障发生的位置和原因。现在常用的方法有专家系统和图论等。研究人员主要研究的都是专家系统。专家系统的基本流程框架如图 2.3 所示。

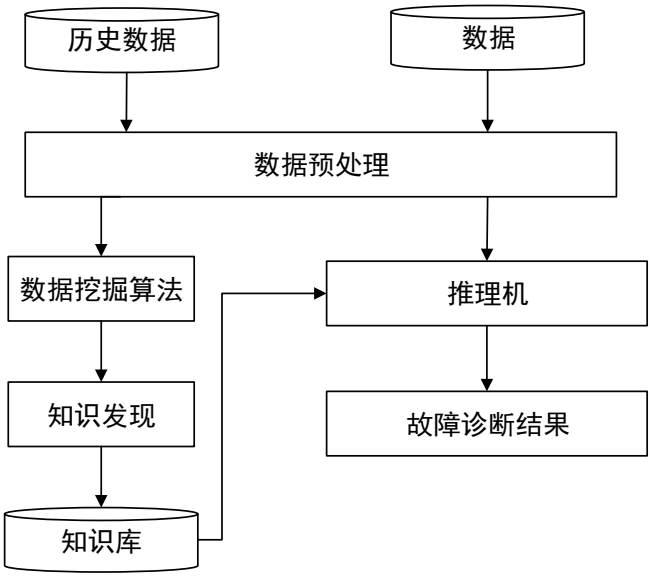

图2.3 专家系统组成图

Feigenbaum 教授认为专家系统是一个智能计算机程序，主要利用知识和推理能力来解决需要专家知识才能解决的问题。专家系统的基本工作原理是首先从采集系统得到数据信息后，利用推理机中保存的各种知识进行相应的推理，快速确定故障。若将一个专家系统当做一个人类专家，其中知识库就相当于人脑子中以前处理过的一些案例，推理机相当于是人的推理过程。所以在建立一个可靠性强、准确度高的专家系统就要保证有充足的数据源，具有优秀的推理算法，这两个方面正是研究人员一直研究的重点。

专家系统的主要特征包含：（1）专家水平的专业知识；（2）能够有效进行推理；（3）知识获取能力；（4）交互能力。其中知识发现能力十分重要，专家系统的基础是对知识的存储和利用，知识获取是得到知识的唯一途径。知识发现就是将已存在的知识从数据源中进行抽取，将其表示成为计算机能够理解的形式，并输入计算机内的转换过程。知识发现一般分为直接获取和间接获取两种不同的方式。直接获取方式的过程是：领域专家向系统提供一定数量的数据及资料，系统通过机器学习将这些数据

和资料按一定的格式整理成知识，形成知识库。间接方式是领域专家将自己的知识用语言及书面的形式整理出来，然后知识工程师在领域专家的帮助下对他们提供的知识进行分析、总结和简化，形成易于被计算机理解的知识表示形式，借助知识编辑工具将知识输入到系统知识库中。知识发现的具体方法也可以分为许多，如机械式、归纳式、解释式、基于神经网络、基于遗传算法等。经过相关专家多年的实践表明，各种知识获取方法都有其局限性，在专家系统开发过程中只能具体问题具体分析，可以将多种方法结合起来进行应用。

知识的表示是为了描述某一事件而做的一组约定，是知识的符号化和形式化的过程。知识表示方法就是将已经获得的各种知识以计算机内部代码的形式加以合理地描述和存储，它的目的是通过知识的有效表示，使专家系统能够利用这些知识做出合理的推理和决策。目前人工智能领域已经出现了多种知识表示方法，如产生式表示法、框架表示法、语义网络表示法、基于粗糙集的表示法等，其中在故障诊断领域应用比较多的方法有产生式表示法。产生式表示法又可以称为规则表示法，它模拟人脑中存在的因果关系，主要是以“IF-THEN”的产生规则的形式来对知识进行表示。即：

$$P \to Q$$

或者

$$\text{IF}\, P \;\; \text{THEN} \;\; Q$$

其中，P代表条件，如前提、原因等；Q代表结果。含义是：若P被满足，则可以得出结论Q。这种方法的优点是：模块化好、表示形式一致，十分符合人类自然的思路。

现在工业中数据源问题已经不是关注的重点，在这些数据源中获取有价值的信息成为了重点，我们将其称为知识发现。传统的知识发现方法，很难满足海量数据下专家系统的要求，所以有必要应用智能算法来进行知识的发现。知识库的建立主要是依靠间接的方法来进行知识发现，主要是由工程师人工建立的，来源于人类专家丰富的经验，然后工程师将专家的经验进行提炼、总结。张荣梅、曹存根等人认为知识发现是基于知识工程方法的瓶颈[27][28]。现在工程师可以利用聚类、关联规则等数据挖掘算法来获取知识，并且能够实时的更新知识库。随着专家系统的运行，知识库会不断的丰富，其分析能力也就不断上升。

2.2.3 基于数据驱动的方法

基于数据驱动的故障诊断方法是近年来的热门研究领域，目前在很多的企业中，设备每天都会产生和存储大量的工况数据，这些数据可分为正常条件下的和特定故障条件下收集到的数据，包含设备的各种信息。

基于数据驱动的故障诊断方法主要的数据源包括两部分：在线设备数据和存储的大量历史离线数据。对于离线数据必须保证其信息的充足和完整，数据能够包含设备

各个方面的情况。基于数据驱动的方法的基本思想是利用获得的设备数据建立历史数据分类器模型，通过所建立的模型来对运行的实时数据进行故障诊断。

2.2.4 数据挖掘与知识发现

数据挖掘(Data Mining,DM)主要是指从大量数据中自动分析并提取人们感兴趣的、隐藏的有价值的知识的过程。它是一门面向应用的交叉学科，里面涉及到人工智能、机器学习、可视化技术和数据库技术等。数据挖掘主要有以下任务：（1）相关性分析；（2）偏差检测；（3）分类和聚类；（4）预测；（5）数据总结等。常用的数据挖掘方法包含有决策树算法、神经网络算法、关联规则算法等。数据挖掘是知识发现的核心部分，是将数据上升到知识的关键步骤。

利用数据挖掘算法建立知识库主要分为五个基本步骤：（1）故障数据库的建立；（2）故障数据与处理率；（3）建立数据挖掘模型；（4）知识表示；（5）知识库的建立。

2.3 本章小结

本章中详细描述了混合模型中的关键步骤，详细介绍了数据预处理、基于知识工程的方法、基于数据驱动的方法以及数据挖掘与知识发现之间的关系。

第三章　基于支持向量机的数据驱动方法

当面对未知的故障时，在应用基于知识工程的方法之前要先利用基于数据驱动的方法来进行诊断，得出诊断结果后，利用数据挖掘算法来进行知识发现，不断地丰富知识库。基于数据驱动的方法是混合模型中十分重要的一部分，在工业大数据环境下，它将会具有很大的发挥空间。

3.1 基于数据驱动的方法简介

在工业大数据环境下，工业设备已经累计的大量的离线数据，而且会产生更多的数据。众多专家都在研究如何通过提取的数据特征并分析其内在规律，实时并有效的将在线数据中包含的故障数据进行检测并完成故障分离，并最终达到保证设备安全。数据驱动方法是指一类不需要研究设备的精确模型，利用设备的海量数据就能够实现对其的故障诊断等目的的技术。

基于数据驱动的故障诊断方法主要有基于统计分析的方法、基于信号分析的方法以及人工智能的分析方法[10]。基于统计分析的方法主要是依靠分析设备数据统计量，从其中变化的提取特征。基于信号分析的方法是利用设备在运行过程中的各种信号分析技术，提取信号时域和频域的特征来确定设备的状态。基于人工智能技术是通过教计算机如何学习、推理和决策等来实现故障诊断。

现阶段，研究基于数据驱动的方法并不是要完全取代传统的方法，由于现阶段整个理论基础、架构体系还不是很成熟。研究该方法只是希望其能够作为传统方法的一个补充，但是研究的目的是希望该方法在理论成熟、架构完善的时候能够完全取代传统的方法。基于数据驱动的方法实际应用过程中，会用到许多的智能算法如神经网络、决策树以及支持向量机等。在实际应用中，由于受到样本数目的限制，一些样本需求量大的算法慢慢开始被人诟病，一直到支持向量机（SVM）被提出。SVM 继承了统计学习理论中的许多优点，现在 SVM 已经在许多数据处理方面得到了应用，如文本分类、图像处理、故障分析等。其核心内容包括：（1）在寻找分类函数时，引入了超平面的概念；（2）在处理非线性样本时，引入了核函数；（3）处理多分类问题。

3.2 支持向量机(SVM)算法理论

SVM 算法是 Vapnik 等人在 20 世纪开始致力于小样本情况下的机器学习研究工作，并建立了统计学习理论的基本体系。支持向量机理论被认为是目前针对小样本问题的最佳理论，它从理论上较系统研究了经验风险最小化原则成立的条件。

3.2.1 SVM 算法在故障诊断中研究现状

在支持向量机(SVM)提出之前，传统的经典统计学大多研究的时候的都默认的是样本量很大的情况，在该情况下研究各种算法的统计性质，如人工神经网络算法需要大量数据训练才能取得比较好的结果。支持向量机算法之所以能够在众多机器学习算法中独树一帜，因为它处理问题的前提不是大样本，而是专门针对小样本而发明的，它能够保证结构风险最小化[29]。

在故障诊断领域，M.Rychetsky 是首次将 SVM 应用于故障诊断中，以发动机为研究对象解决爆震现象，利用 SVM 实现了对发动机爆震现象的检测[30]。SVM 算法正式在故障诊断领域发挥作用是 2001 年 University of Liverpool 大学的 L.B.Jack，A.K.Nandi 等人与肖健华将其 SVM 算法成功运用到轴承和电机的故障检测中。国内胡寿松教授应用支持向量机算法来对数据的残差进行分类，主要利用了该算法在小样本的情况下依然具有很好的分类效果，应用对象是歼击机，对其进行故障隔离[31]。2005 年左右 SVM 在故障诊断领域得到的大量专家的重视，随之便得到了快速的发展。上海交通大学 GanyunLv 等对 SVM 实际应用中的数据预处理、分类扩展等问题进行了研究，对变压器进行了故障诊断[32]。由这些国内外专家的研究发现，这些研究涉及到 SVM 的方方面面，从数据预处理到最终的应用层面。SVM 算法应用十分广泛，不仅仅在民用机械上，在大量的军工装备，航空航天等专业领域也能够看到该算法的身影，现在依然有好多专家在研究如何对其进行改善，使其功能更加强大。

3.2.2 支持向量机（SVM）算法理论

SVM 算法是机器学习算法中的一种，机器学习是指计算机在完成一次计算任务后，积累本次任务的经验，并利用这些经验不断改进自己解决问题的能力。其模型如图 3.1 所示：

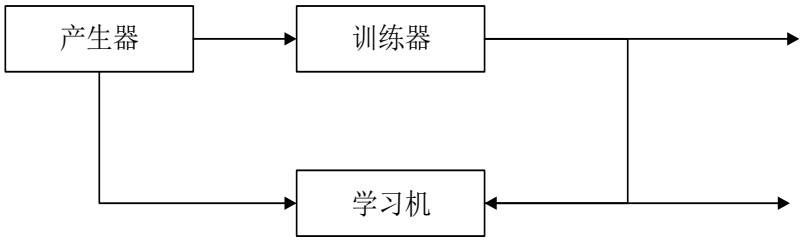

图3.1 机器学习模型

为了解支持向量机的原理，必须首先介绍 VC 维概念。VC 维的定义是：假设有 X 个样本，若某分类器将 X 个样本用 2^x 种不同的方式分隔开，该分类器的 VC 维就是样本个数 X 。简单的理解 VC 维就是指某个分类器能够实现最多分类的样本个数，常用其来判断某分类器和样本的复杂度。在实际应用中，一般应用结构风险最小化（SRM）

来确保 VC 维的大小满足要求。

经验风险最小化(ERM)：

$$R_{emp}(\omega)=\frac{1}{n}\sum_{i=1}^{N}L[y_i, f(x,w)] \tag{3-1}$$

$L[y_i, f(x,w)]$ 代表 y 的损失集，$\{f(x,w)\}$ 是指预测函数，N 为样本数，ω 为广义参数。

它们之间满足：

$$R(\omega)\le R_{emp}(\omega)+\phi(n/h) \tag{3-2}$$

其中：$\phi(n/h)=\sqrt{\frac{h(\ln(2n/h)+1)-\ln(\eta/4)}{n}}$，$R(\omega)$ 是实际风险。

在 SRM 原则下，分类器在设计的时候就必须保证所选择的算法既能够保证分类的效果最优，又必须保证 SRM 的值最小，对算法提出了进一步的要求。

支持向量机算法是从线性可分情况下的最优分类面提出来的，如图 3.2 所示就是一个线性分类图，将两类样本分开的实线叫做分类超平面。SVM 算法的目的就是要找到一个效果最好、性能最优的超平面。

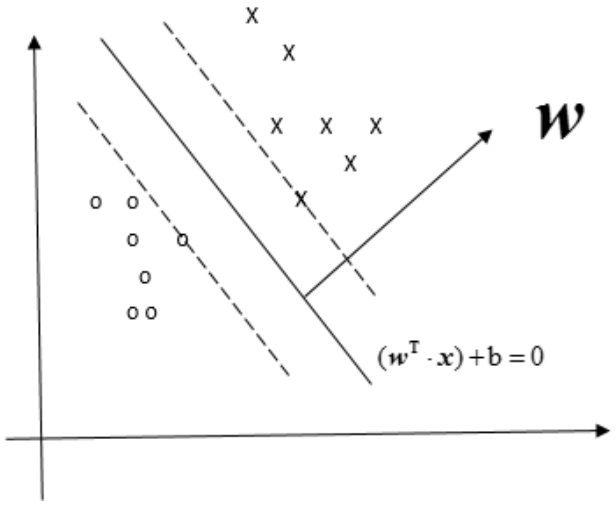

图3.2 支持向量机分类图

先介绍线性可分的样本情况，假设训练样本集为

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

将超平面以及它的约束定义为：

$$\begin{aligned}&(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x})+\mathrm{b}=0\\&s.t.\ \ y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+\mathrm{b})\geq 1, i=1,2,...,\mathrm{n}\end{aligned} \tag{3-3}$$

定义(w,b)关于训练数据集T的函数间隔为超平面(w,b)关于T中所有样本点(x_i,y_i)的函数间隔最小值，其中x是特征，y是结果标签，i表示第i个样本。

我们定义函数间隔为：

$$\hat{\gamma}=y(w^T x+b)=yf(x) \tag{3-4}$$

将函数最小间隔定义为：

$$\hat{\gamma}=\min\hat{\gamma}_i, i=1,2,\cdots,n \tag{3-5}$$

如果成比例的改变w和b，虽然超平面没有改变，但分割函数的值却发生了变化，于是引入真正意义的点到超平面的距离——几何间隔。

几何间隔的定义为：

$$\tilde{\gamma}=\frac{\hat{\gamma}}{\|w\|} \tag{3-6}$$

将函数间隔定义为1，这个改变并不会对超平面有任何的影响，但是可以将整个计算变的简单，于是优化目标就变为：

$$\begin{aligned}&\max\frac{1}{\|w\|}\\&s.t.\ \ y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+\mathrm{b})\geq 1, i=1,2,...,\mathrm{n}\end{aligned} \tag{3-7}$$

将最大化问题转化了求最小值的问题，方便求解：

$$\begin{aligned}&\min\frac{1}{2}\|w\|^2\\&s.t.\ \ y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+\mathrm{b})\geq 1, i=1,2,...,\mathrm{n}\end{aligned} \tag{3-8}$$

应用 Lagrange 对偶性，为每一个约束条件加上一个 Lagrange 对偶变量α：

$$L(\boldsymbol{w},b,\alpha)=\frac{1}{2}\|\boldsymbol{w}\|^{2}-\sum_{i=1}^{n}\alpha_i(y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+\mathrm{b})-1) \tag{3-9}$$

考虑约束条件，将所求问题转化为：

$$\min_{w,b}\max_{\alpha_i\geq 0}L(w,b,\alpha) \tag{3-10}$$

由于求解最小值问题不好求解，故转化为求解最大值问题：

$$\max_{\alpha_i\geq 0}\min_{w,b}L(w,b,\alpha) \tag{3-11}$$

首先要让$L(w,b,\alpha)$关于w,b最小化，然后求α的极大。分别对w,b求偏导数，计算最优的值$\boldsymbol{w}^{*}$和b^{*}。其中：

$$\begin{aligned}\boldsymbol{w}&=\sum_{j=1}^{n}a_j y_j x_j\\ b&=-\frac{1}{2}(\max_{i:y_i=-1}\boldsymbol{w}^{T}\boldsymbol{x}_i+\min_{i:y_i=1}\boldsymbol{w}^{T}\boldsymbol{x}_i)\end{aligned} \tag{3-12}$$

利用公式求解对偶变量α的最优解：

$$\begin{aligned}&\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^{T}x_j\\ &s.t.\ \alpha_i\geq 0, i=1,2,...,n\\ &\sum_{i=1}^{n}\alpha_i y_i=0\end{aligned} \tag{3-13}$$

解得最优解$\boldsymbol{a}^{*}=(a_1^{*},a_2^{*},...,a_n^{*})^{T}$

但是在面对真实的设备数据时候，肯定是非线性的居多，则线性分类就没有了用武之地，或者说就很难发挥作用。而且其实非线性问题不仅仅在支持向量机时候会遇到，在其他算法应用的时候也经常碰到，所以科研人员就提出了核函数技术。

核函数(Kernel Function)是指向量在隐式映射后空间中的内积函数。根据泛函理论的 Mercer 定理[33]，利用核函数可以直接在原空间中计算。若按照传统的方法来处理非线性问题，会发生维灾难。

如待求解的问题如下式所示：

$$\begin{aligned}&\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j<\phi(x_i),\phi(x_j)>\\&s.t.\ \alpha_i\geq 0, i=1,2,...,n\\&\sum_{i=1}^{n}\alpha_i y_i=0\end{aligned}\tag{3-14}$$

其中$<\phi(x_i),\phi(x_j)>$就是高维空间，利用核函数技术能问题转化为原空间内的问题：

$$\begin{aligned}&\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j)\\&s.t.\ \alpha_i\geq 0, i=1,2,...,n\\&\sum_{i=1}^{n}\alpha_i y_i=0\end{aligned}\tag{3-15}$$

其中$K(x_i,x_j)$是核函数。

使用了核函数，也有一些很难处理的情况，例如数据有噪音，数据中有离群点，离群点是指那些不在正常位置上的点。通过对 SVM 原理的研究，这些离群点很可能会对分类造成很大的影响，因为超平面本身是由少数几个支持向量构建的，如果这些支持向量中存在有离群点的话，对分类的影响就会很大。

若在优化的时候考虑离群点，则优化的函数变为：

$$\begin{aligned}&\min\frac{1}{2}\|\boldsymbol{w}\|^2+C\sum_{i=1}^{n}\xi_i\\&s.t.\ y_i(\boldsymbol{w}^{\mathrm{T}}\cdot\boldsymbol{x}_i+\mathrm{b})\geq 1-\xi_i, i=1,2,...,\mathrm{n}\\&\xi_i\geq 0, i=1,2,...,\mathrm{n}\end{aligned}\tag{3-16}$$

其中ξ_i（$i=1,2,...,\mathrm{n}$）称为松弛因子，它是指数据点$\boldsymbol{x}_i$允许偏离的函数间隔的量。

经过变化后，可以转化为如下对偶问题：

$$\begin{aligned}&\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j\\&s.t.\ 0\leq\alpha_i\leq C, i=1,2,...,n\\&\sum_{i=1}^{n}\alpha_i y_i=0\end{aligned}\tag{3-17}$$

该算法主要优点有：（1）它是能够解决维灾难等问题；（2）有很宽的作用领域；（3）SVM 对于非线性数据的处理能力和预测能力会随着实验数据的完备不断改进；（4）与其他的机器学习算法比较，SVM 训练起来比较简单。

3.3 核函数及其参数的选择

核函数是由 Aizermann 等人引入机器学习领域中，1992 年 Vapnik 等利用该技术成功将线性 SVM 推广到非线性 SVM 中。在模式识别理论中，若某个数据集在低维空间线性不可分，则可以通过非线性映射到高维空间中来实现线性可分，但是如果直接采用这种技术在高维空间中进行分类，则存在确定非线性映射函数的形式和参数、特征空间的维度等问题，而最大的障碍是特征空间运算时候存在“维灾难”，采用核函数技术能够有效解决该问题。

核函数有很多种，并不是任何一个核函数都能够将非线性数据集分类开来。一个分类器处理非线性问题的能力与核函数的选择有很大关系，不同形式的核函数其对非线性样本的映射效果不同[34]。

设 $x_i, x_j \in X$ ， X 属于 $R(n)$ ，函数 Φ 为非线性函数，则：

$$K(x_i, x_j) = <\Phi(x_i), \Phi(x_j)> \tag{3-18}$$

$<,>$ 为内积， $K(x_i, x_j)$ 为核函数。

支持向量机、PCA 和 Fisher 判别法等在其建立分类器的过程中都需要用到核函数。核函数主要特点有：（1）计算量将不会再受特征空间维数的限制；（2）无需关心变换函数；（3）核函数不同那么所确定的非线性映射也是不同的[35]。核函数有以下几个：

（1） 线性核函数

$$K(x_i, x_j) = x_i \cdot x_j \tag{3-19}$$

（2） 多项式核函数

$$K(x_i, x_j) = [(x_i \cdot x_j) + \mathrm{c}]^q \tag{3-20}$$

齐次多项式核：

$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^q \tag{3-21}$$

非齐次多项式核：

$$K(x_i, x_j) = (x_i \cdot x_j)^q \tag{3-22}$$

（3） 高斯核函数(RBF 核)

$$K(x_i, x_j) = \exp\{-\frac{\left|x_i - x_j\right|^2}{\sigma^2}\} \tag{3-23}$$

（4） Sigmoid 核函数

$$K(x_i, x_j) = \tanh(b(x_i \cdot x_j) - \mathrm{c}) \tag{3-24}$$

线性核函数主要是应用在数据集是线性可分的，经过与其它几个核函数对比可以发现，线性核函数参数很少，函数比较简单，所以在训练和执行的时候速度相对较快。在对多项式核函数处理的时候，由于它是一个全局核函数，所以求得的解就是全局最优，它泛化能力强，但学习能力较差。若处理的数据集没有任何其他的先验知识，一般选 RBF 核。RBF 核能够保证其函数运行结果的范围在 0~1 之间，经过变换后可以简化计算。变量σ在 RBF 核性能中起着关键作用，σ称为尺度参数。σ的大小可以分类讨论：（1）σ等于 0，或者在 0 的附近，样本都是支持向量，无意义；（2）$\sigma > 0$，且σ的值趋近无限大，那么就会将所有的样本分到一个类中，不具备分类能力；（3）$\sigma > 0$，但是值比较小，分类准确率会很高，出现“过学习”现象。Sigmoid 核函数与其他三个不同，因为它并不是在任何情况下都满足非负定的要求，若希望其能满足要求，就必须对参数b和c进行限制，这两个参数必须要满足$b > 0, c < 0$。Sigmoid 核函求解的是凸问题，得到的解是全局最优解。对上述四个核函数的研究后，有专家就希望能够建立混合核函数，经过研究发现，Sigmoid 核函数与高斯核函数结合起来得到的效果最优。

核函数的选择方法主要有：（1）利用以前发生过的案例，看当时核函数的参数选择，也就是利用历史经验来预先选择核函数；（2）在具体的案例中，对所有有可能的核函数一一进行实验，然后分析核函数处理的结果，将误差最小、综合效果最好的定的要选择的核函数；(3)取长补短，将两个或多个核函数结合起来，形成混合核函数，

可能会得到较好的特性。

核函数方法的实施具体如下：

（1）获取数据源，利用最为方便的手段来，视具体情况而定。采集并保存样本数据，进行预处理；

（2）选择效果最优的核函数；

（3）将样本映射到特征空间；

（4）在特征空间进行分类。

有人将核函数称为“非线性革命”，可见核函数在分类器中的作用。在“发明”了核函数后，专家们就开始研究如何将核函数中的参数调整到最优，达到该核函数的最好效果。

一个核函数中参数的确定方法有很多，经验确定法、交叉验证法和梯度下降法等。经验确定法由于其简单可行，所以应用最为广泛，需要专家人士。梯度下降法出现的比较早，基本思想是先选择一个初值赋值给参数和步长，主要是应用迭代的方法来达到要求，可以选择分类的错误率作为标准，来判断是否停止迭代，该方法中初值的选择很关键。交叉验证法是将参数的可能取值区间均分为n等份，在每一个区间内随机选取参数，取每个区间中分类错误率的均值。利用这些均值得到一个综合错误率即参数在总区间中的错误率。最后将划分的几个区间范围和对应的数据合并，重新选择参数。

3.4 多分类问题

支持向量机的产生是为了解决两类分类问题，但是在设备在运行过程中，经常会有很多故障同时发生，必须要面对多分类问题。如液压系统中，若以马达的压力和转速为监测对象，通过收集到的数据发现压力和转速数据异常，可能是由于油缸漏油，也可能是因为油液中含有气体，也可能是马达自身出现了问题。所以当设备出现故障，可能会有多种故障源，判断时候要考虑多分类问题，尤其是复杂设备。

SVM 解决多分类问题的方法主要有两种形式：一种分类方式是在原有的支持向量机的优化函数上进行改动，引入多类机制，从而解决多分类问题。该方法由于目标函数过于复杂，实现困难大，不适合在实际中进行应用；另一种方法是构建多个两类支持向量机并以一定的结构结合起来形成一个多类分类器，间接解决多类分类问题，这类方法应用的比较多，常用的两种方法是全局多类支持向量机和组合多类支持向量机方法[36]。

3.4.1 全局多类支持向量机

该方法是在数据集的全局，利用算法求得一个最优解。算法实现如下：

训练集为$\{(x_i, y_i),...(x_m, y_m)\}$，$x_i \in R^n$，$y_i \in (1,2,...\mathrm{M}), i=1,2,...,m$。能够找出判别式$f(x)$使得所有$x$都能找到$y$值与之对应。首先构造 n 个二分类器，$\mathrm{sgn}[f_k(x_i)]$根据向量是否属于相对应类别来判断，如果向量$x_i$是属于第$k$类，则$\mathrm{sgn}[f_k(x_i)]$值为 1，否则为-1。经过前面的处理后，选择函数$f_k(x), k=1,2,...,n$所求值中最大值，找出其最大值所对应的类：$m=\arg\max\{f_1(x_i),..., f_n(x_i)\}$，通过该方法需要构造出 n 类分类器，才能够实现多分类。

3.4.2 组合多分类

在组合多分类中，包含两个方法：一对一方法，一对多方法[37]。

（1）一对多方法

先构造一个分类器，每个分类器都是二分类。然后根据第一个分类器的分类结果，从样本中选择出一类，将其认为是正训练样本，那么自然其它样本就为负训练样本。然后再继续在负的样本中，同样应用二分类器，在其中挑一类样本作为正样本，依次循环下去，直到将所有的样本分开。由此，对于k类分类问题，就必须形成k个二分类器才能将所有的样本分开。

一对多方法分类速度快。但是当处理的数据过多时，训练速度会明显减慢。而且可能存在不可分现象，若某一类样本与其他样本特征差别较大时，由于训练样本不平衡会影响分类精度。

（2）一对一方法

该方法是与一对多不同，直接在两个不同类之间构造分类器。所以在解决n类问题时，需要建立$n(n-1)/2$个分类器。在分类的时候，肯定会有重叠部分，所以在最后确定样本属于哪个类时候，使用投票法，考虑所有二分类器，样本归属于得票最多的类。很明显，当数据量增加时候，分类的运算量急剧增加，分类速度会变的很慢。

3.5 本章小结

首先介绍了基于数据驱动方法的思想，即利用设备运行时候的工况数据来对设备的状态进行分析。然后详细介绍 SVM 的基本原理、核函数及参数选择问题、多分类问题等。

第四章　利用数据挖掘技术进行知识发现

当核函数以及分类函数比较复杂，在进行故障判断时候，消耗的时间就会较长。基于知识工程的方法直接利用知识库中的规则，执行效率快，判断准确度高，即使在数据量很大的时候，依然能有很高的准确率。

4.1 基于知识工程的方法

基于知识工程的方法出现的比较晚，但是现在不仅得到了科学界的高度重视，在工业中也有广泛的应用。根据故障的先验知识来建立设备的定性模型，这些先验知识包括设备的故障集合、故障与现象之间的关系等。一旦有故障发生，就按照提前设计的算法进行运算或搜索，判断出故障的具体信息。现阶段，主要应用的方法有专家系统和图论，本文选择应用专家系统。

4.1.1　专家系统研究现状

专家系统是为了给特定领域提供“专家”级服务的程序，它内部包含了领域内大量的知识经验。它能够模拟人类思维来进行决策支持，所用的算法越是合理，得到的决策就越正确。一般一个专家系统要想具有较高智能化，就必须具备自己获取知识的能力、具备相当高的推理能力。

斯坦福大学 E.A.Feigenbaum 教授在 1965 年提出了化学领域的专家系统 DENDRAL，它的解决问题的能力十分强大，随着该专家系统的多年运行，现如今已经能够达到专家水平。在故障诊断方向，专家系统开始阶段主要应用在船舶和机床的振动诊断和证据分析中[38]。国内吴海桥等人将专家系统应用于大型客机的故障诊断中，在他的博士论文中首次提出了专家系统的开发生存期模型，这一点比较具有创新性。他将故障分类为常见型和偶发型两类故障。对于常见型应用事例推理法，而对于偶然型则应用基于模型的推理方法。金亮亮等针对航天器中的姿态控制器系统，将基于规则方法和故障树结合起来，建立了针对航天器的专家系统[39]。樊彬彬等人研究了专家系统中综合推理机在盾构机故障诊断中的应用，将基于不确定性、规则、实时参数三种情况进行了结合[40]。

经过对相关文献的查阅和统计，如今专家系统主要的研究方向有以下几个方面：（1）利用模糊逻辑来表示和处理知识；（2）对专家系统中的匹配算法进行研究；（3）专家系统在故障诊断方面的应用。由于专家系统自身特有的知识结构和推理机制，非常适合建立一个故障排查系统。

4.1.2 专家系统简介

研究专家系统的目的就是希望能够将某个领域中的专家经验进行收集和总结，然后利用这些经验来实现对故障的快速诊断。判断某专家系统的性能主要依据是：（1）所设计的专家系统中所具备的知识量，是否包含有充分、大量的领域知识；（2）专家系统对这些知识的应用能力。这两点缺少任何一个都不能称之为一个好的专家系统。

专家系统由下面几个模块构成：（1）知识库：知识库的主要作用是存储知识和案例的，专家总结的经验知识和历史经验都存储在知识库中；（2）推理机：根据当前知识库中的知识来导出结论；（3）知识发现模块：该模块是专家系统从知识库中发现新的知识，不断丰富知识库，增强专家系统的处理能力；（4）知识表示：将知识表示成统一格式，表示成易于处理的格式；（5）人机交互：这是操作人员与计算机机器之间传递信息的接口。

专家系统研究的主要问题在于怎样利用计算机，发现并自动获取。传统的专家系统中是通过人工来获取的。现在研究如何将数据挖掘方法应用来进行知识发现。它主要从海量设备相关历史数据中发现知识，这些历史数据主要包含的是设备在过去一段时间内运行的工况数据以及与这些数据相关的设备的状态数据，将这些数据作为挖掘的数据源，计算机通过对这些数据的学习来总结出一些规则。对于特定的设备，得到的这些规则是对于这类设备是通用的，能够解决该设备的某一类问题。

4.1.3 专家系统中知识发现

知识发现就是指科研人员自己人工或利用智能算法在外部数据源中发现知识，并将这些知识转化到计算机中的过程[41]。知识发现一直是专家系统的“瓶颈”，知识发现是从实际案例中来发现规则或者直接从专家处输入规则，慢慢来将丰富知识库，不断完善专家系统的能力。

专家系统研究初期，知识发现是通过人工来实现的，都是通过一些专家对历史记录的总结。通过工程师来发现知识，然后对这些知识进行修改，将其变换成具有统一格式的规则，而后存入到知识库中。人工处理方法速度十分的缓慢，效率十分低下，仍然必须依赖专家。随着研究数据挖掘、分析的研究人员不断增加，发现知识的方法也发生了变化，研究人员希望能够自动发现知识。这种方法不需要专家直接同系统对话，而且也不需要计算机工程师的干预，将发现知识导入知识库中。自动式知识发现方法虽然才刚刚起步，但却有十分光明的前景。

4.2 利用关联规则算法进行知识发现

4.2.1 数据挖掘与知识发现

传统的专家系统会将信息程序化，归纳好，建立知识库，完全依赖的是书本上的知识和领域专家手工来进行知识库的修改、扩充等。解决知识自动获取这一难题，数据挖掘成为了研究的热点，数据挖掘是一个多领域多技术合作的过程，也是一个不断优化问题解决方案的过程[42]。

利用数据挖掘算法来获取知识虽然该算法还在研究的初期，但是由于它不需要专家的参与，是完全自主的，所以具有很大的发展空间。数据挖掘一般应用机器学习来进行知识发现，学习过程总与环境、知识库有关，环境和知识库是某种形式的信息集合，分别代表外界信息源和系统具有的知识。

数据挖掘的流程如下图4.1所示：

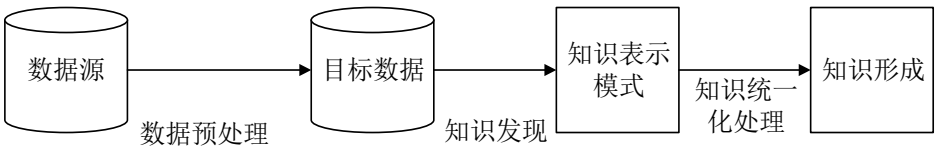

图4.1 数据挖掘流程图

4.2.2 关联规则算法 Apriori

关联规则挖掘是数据库中知识发现的一个重要分支，是数据挖掘中一种简单实用的规则。Apriori 算法是关联规则的典型算法，该算法主要应用的是逐层搜索迭代方法，在遍历(k+1)-项集时应用其前一项 k-项集。一旦规则被产生，将那些大于用户给定的最小支持度的规则保留下来，作为候选项集。

该算法主要有三点比较关键：（1）从频繁1-项集开始，按照逐层迭代法，只遍历一层；（2）在确定频繁项集的时候，使用的是先产生再测试的方法；（3）遍历结束后，通过计算支持度、置信度来确定规则。算法总迭代次数是通过频繁项集中的最大长度来决定的，若最大频繁项集长度为$k_{\max}$，则最大迭代次数为$k_{\max}+1$。

支持度和置信度分别定义为：

$$s(X \to Y) = \frac{\sigma(X \cup Y)}{N} \tag{4-1}$$

$$c(X \to Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{4-2}$$

Apriori算法步骤：（1）首先对整个数据集进行单遍扫描，计算出每项的支持度；（2）产生k-项集。由于已经得到了1-项集，利用逐层迭代算法就能确定出k-项集；（3）确定候选项集。最后一次扫描数据集，计算支持度。通过这一步处理，将删掉不满足要求的候选项集。

由于算法的复杂度对整个知识发现的处理效率有很大的影响，所以必须考虑其复杂度。影响算法复杂度的因素有：（1）支持持度的阈值大小。支持度的阈值大小十分关键，若阈值设定过小，就会产生很多的频繁项集，大大增加了计算的复杂度。但是若规定的阈值较大，有些规则就会漏掉。朱习军等人对该问题进行研究提出了一种估算方法，该方法了利用支持度的历史值来估算当前值，然后应用牛顿插值公式来自动获取支持度阈值[43]；（2）数据的维度，也就是项数对算法复杂度也有十分重要的影响。数据的维数越多，自然其产生的频繁项集也就越多，处理就越复杂。处理的数据维数越多的话，消耗计算机的存储空间就越多。

4.2.3 Apriori 算法中规则发现

规则是在产生的频繁项集中得到的，但并不是所有的频繁项集都能够成为规则，只有那些满足了置信度阈值的频繁项才能成为规则，将这种方法称为基于置信度的减枝。若例如某频繁项集中有k项，就能够产生2^k-2个关联规则。

Apriori算法产生关联规则也是通过逐层方法来产生的。规则生成的步骤为:（1）将频繁项分为前后件，提取后件数目为1个的规则，找出其中置信度满足要求的项；（2）利用上一步产生的规则集再继续产生候选规则。例如一个频繁项为$\{a,b,c,d\}$，若$\{acd\}\rightarrow\{b\}$和$\{abd\}\rightarrow\{c\}$是满足要求的后件为1，且满足置信度的规则，则可以得到候选规则$\{ad\}\rightarrow\{bc\}$，如图4.2所示。如果规则$\{bcd\}\rightarrow\{a\}$的置信度没有达到规定的阈值，则删除掉所有包含属性a的规则，包括$\{cd\}\rightarrow\{ab\}$，$\{bd\}\rightarrow\{ac\}$，$\{bc\}\rightarrow\{ad\}$和$\{d\}\rightarrow\{abc\}$。图4.2中黑色部分在规则形成过程中会被删除掉，是不满足要求的规则。

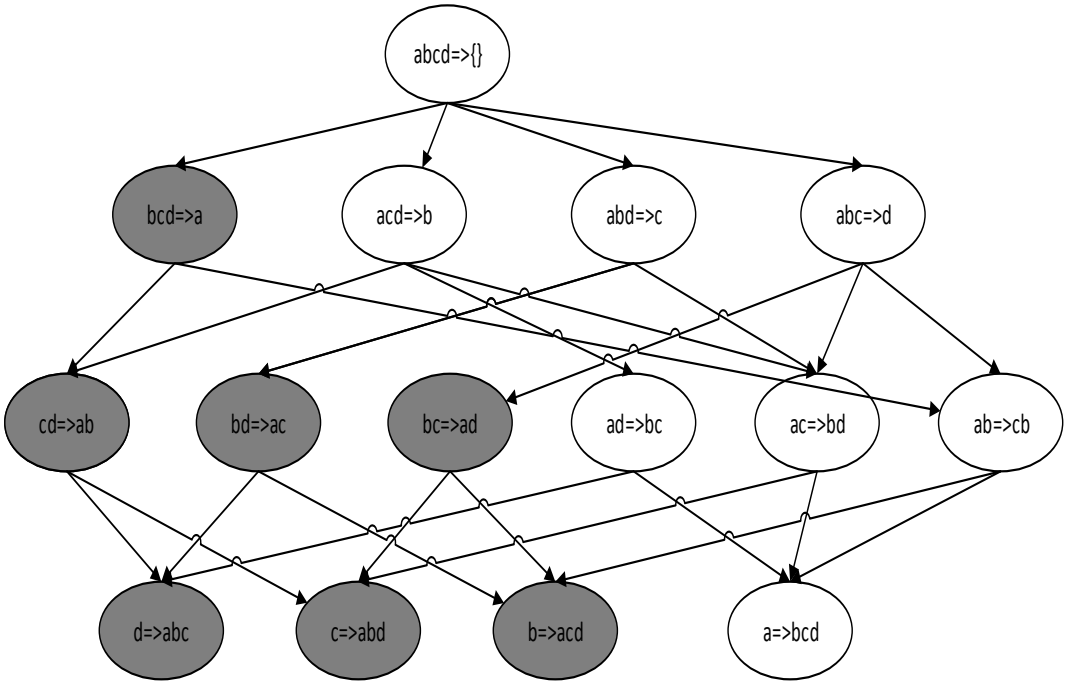

图4.2 规则生成图

4.3 利用聚类算法进行知识发现

聚类算法是根据样本的属性来将未分类的样本进行类别的识别，将类似样本归为一类的算法[44]。在许多的商业中，已经能够看见聚类算法应用的影子，例如在超市中聚类算法能帮助分析人员判断出不同消费群体的习惯。聚类分析的主要应用包含模式识别，数据分析，图像处理以及市场研究等。聚类算法主要是根据描述对象的属性值计算得出的相似度，将数据对象分组成为多个类或簇，在同一个簇中的对象之间具有较高的相似度。聚类主要有两个部分组成：（1）发现簇；（2）对簇的合理解释。

设备在运行时候收集到的数据大多数都是数值形式，如压力、转速、振动等信号。设备不同的状态下，收集到的工况数据肯定是不同的，通过对这些数据进行聚类，可以得到不同状态的类群。聚类算法的中心点的大小以及距离中心点的距离可以作为一个阈值规则，对新到达的数据进行判断。

4.3.1 k-means 算法简介

聚类算法主要是对每一种故障情况进行聚类。对与每一类故障的样本集，通过聚类算法，将最佳聚类阈值存储到阈值表中。

k-means 算法将 n 个样本数据分为 k 个聚类，更重要的是要保证每个聚类中的样本数据到该聚类中心的平方和最小。

假设训练样本为：$\{x^{(1)},...,x^{(m)}\}$，每个 $x^{(i)} \in R^n$。步骤如下：（1）随机选取 k 个点，分别定义为 $\mu_1,\mu_2,...,\mu_k \in R^n$；（2）利用公式来计算质心，达到规定次数或者运算到收

敛为止。

对于每一个样本 i，通过公示来计算其应该属于的类：

$$c^{(i)} := \arg\min \left\| x_j^{(i)} - \mu_j \right\|^2 \tag{4-3}$$

对于每一个类 j，要不断地重复计算该类的质心：

$$\mu_j := \frac{\sum_{i=1}^{m} 1\{c^i = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c^i = j\}} \tag{4-4}$$

k-means 步骤图 4.3 所示：

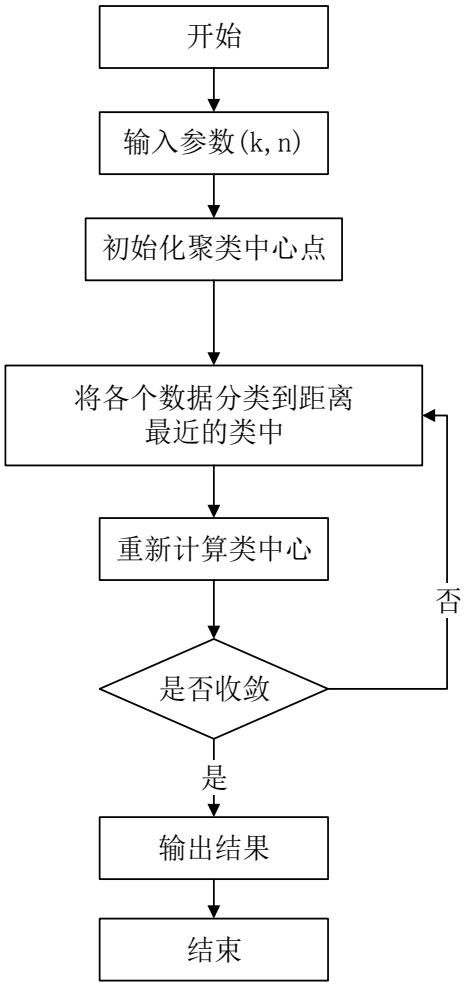

图4.3 k-means 算法流程图

对于 k-means 算法的结束收敛条件，我们定义畸变函数如下：

$$J(\mathrm{c},\mu)=\sum_{i=1}^{m}\left\|x^{(i)}-\mu_{c}^{(i)}\right\| \tag{4-5}$$

J 是一个非凸函数，在选择质心初始位置时候就要慎重。根据研究人员对 k-means 算法的实例研究，算法在达到局部最优就能够满足要求。若想要解决解决局部最优这个问题，可以应用一个简单的方法，多选几次不同的初始值，然后找出其中最小的 J 值以及最小 J 值对应的 μ 和 c 。

4.3.2 k-means 算法进行知识发现

在本文中上述提到的关联规则算法是基于规则的推理方法来实现知识发现的，其形成的知识库中主要是由许多规则组成，在设备诊断过程中，是通过模式匹配来完成的。聚类算法的实质是将设备的数据集分成多个相似的群簇，若进行处理的对象对精度要求不是很高或者数据量相对很大的时候，可以将这一个群簇中的对象作为一个整体来处理。

聚类计算在通过多次运算后，能够确定对应的最佳聚类结果，可以将该结果作为知识库中的知识。而且在复杂的设备在运行时候，传感器收集到的数据大多数都是数值型的数据，我们可以通过聚类算法得出不同状态下的数据均值等特征信息，利用这些信息可以判断出现在设备处于什么样的状态。现在大多数的关联规则算法都无法直接对数值型的数据进行判断，一般会将数值型数据转化为某个字符或者是一段区间表示。

4.4 本章小结

首先介绍了基于知识工程方法中的专家系统，然后介绍了如何利用大数据挖掘和分析技术来进行知识发现，分别介绍了关联规则算法和聚类算法在知识发现中的应用。对于应用到的算法关联规则算法 Apriori，详细介绍了其工作原理以及知识规则获取方法。最后介绍了 k-means 算法，通过聚类算法来实现知识获取。在设备运行时，收集到的数据大都是数值型，聚类算法具有一定的优势。

第五章　盾构机管片拼装系统故障诊断

有专家将盾构机称之为“国之重器”，但是根据对国内正在使用的盾构机了解，大部分都是引进国外的设备。一般会应用德国、日本的产品，这些设备造价都十分昂贵。盾构机在施工过程中，只能前进不能后退，所以一定要保证其在施工过程中的健康运行，对于盾构机的故障诊断研究十分重要。

5.1 盾构机简介

盾构机分很多种类，本文研究的是土压式盾构机，构造十分复杂，主要是对土体进行切削，在城市的地铁施工中一般都应用该类盾构机。土压式盾构机的主要组成部件有：盾构主体、切削的刀盘、排土装置、铰接装置和管片拼装[45]。由于本文主要研究的是管片拼装机的液压系统，管片拼装机主要负责在盾尾内将管片按拼装成管环。

盾构机是十分庞大，结构复杂，涉及到多个学科，难以通过某一个单独现象就判断故障是否发生以及故障发生的位置。在施工时候，需要检测的量多达数十个，而且每一个参数和另外的参数之间存在一定的联系，建立物理或数学模型十分困难，所以研究应用基于混合故障诊断模型具有一定的价值。在整个盾构机设备中，其主要驱动是依靠液压来进行的，液压系统也是最容易发生故障的部位，一般的故障包含：漏油、油液污染等。

管片拼装是施工中的一个重要环节，能够防止地表沉降，在施工建环阶段片建环一定要保证其高效、可靠[46]，管片一般应用$3A+2B+1K$模式，如图5.1所示。本文选择管片周向运动液压系统为研究对象。在施工过程中要完成管片的精确定位，管片必须能够在6个方向上都能够自由运动：轴向直线运动、径向直线运动、周向回转运动、以及管片姿态微调的3个运动，其中周向回转运动采用两台高速液压马达，经过星齿轮减速机减速，小齿轮传动回转支撑的大齿圈，带动整个管片拼装机旋转。

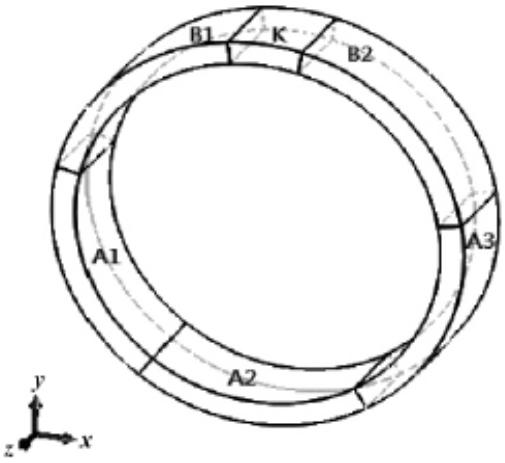

图5.1 $3A+2B+1K$ 管片图

5.2 液压系统常见故障简介

本文以液压系统为研究对象，分析设备在不同状态下液压系统的数据，对液压系统进行故障诊断，研究基于数据驱动方法中算法。故障诊断的基本思想为：假设状态空间S是被检测对象全部有可能发生的状态，S中的状态可以分为两种或多种。所有的特征构成特征空间Y，Y是由可观测量的取值范围构成的。Y与S有关系，当系统处在某一个S中，就能确定对应Y。故障诊断的是模式识别的一种，将待判断状态模式与历史样本进行对比，利用建立的分类器来决定待检模式应该划分为哪一类模式。首先必须建立分类器，通过提取系统的状态S和相应的特征Y进行训练，最重要的是将样本规模降低为适用于支持向量机的小样本模式。

利用支持向量机算法来进行故障诊断，在国内外都有很多专家在研究。杨琦等人将SVM算法应用在液压系统中，针对液压设备中的克令吊为研究对象，实现对液压系统多故障诊断[47]。王盈等人研究了双层模糊支持向量机算法，分析液压系统的运行机理，建立了一个故障诊断模型，并使用该摸型对液压系统进行故障诊断，应用了模糊SVM。针对收集到数据集中存在模糊性的缺点，在其论文中提出了一个双层网络模型的故障诊断方法[48]。

在设备运行过程中，液压系统的油液污染和漏油等是经常发生的故障，主要是由于泵在施工的时候磨损或者在工作时候会油缸里进入空气。液压系统工作中主要的参数有流量、压力、温度、泄漏量以及液压马达的速度等，经过对这些参数的测量和分析，能够实现对夜压系统的故障诊断。

液压系统在运行过程中发生的故障可能单独发生，也可能同时发生[49]。对液压系统进行故障诊断时，首先对系统的工况数据进行采集，然后从这些工况数据提取特征

信号，找出故障与这些信号的对应关系，进而实现来对系统的诊断。因此，从系统众多信号中寻找这些能够区分故障状态的信号就显的十分重要。目前研究发现，通过对油液进行分析得出油液中各成分变化趋势，能够实现对故障的诊断和预测，也可以根据温度来实现对故障的粗略判断，从实际温度曲线与正常温比较，就能够发现液压系统的运行状态。在进行数据采集的时候，温度变量受到的外界干扰比较多，而且由于温度具有滞后效应，再加上外界环境等因素的影响，一般在进行判断时候，温度只作为一个辅助量，从来不作为主要的参考变量，主要是对故障诊断其他参数做一定的补充。现在的液压系统基本上都是基于帕斯卡定律的静态液压传动系统，因此通过一些数据分析软件能够利用压力判断出设备中液压系统的工作状态。经过实践发现，油液的压力能够直接反映出系统的工作状态，压力可以作为状态的载体，在测量的时候，信号容易测量，而且灵敏度度高。目前，在对设备状态检测应用中，以压力作为特征信号的研究较多。

本文参考文献[46]利用AMESim液压仿真软件来对该系统中的周向回转液压系统进行仿真建模，通过修改模型中的参数来模拟液压系统中的故障。本文通过改变仿真参数来模拟液压系统的三种故障模式，模拟的故障分别是油液污染、漏油以及正常情况。

5.3 周向回转液压系统仿真

5.3.1 AMESim 软件简介

AMESim是一种基于直观图形界面的仿真平台软件，仿真平台和MATLAB不同，该平台可以直接对设备的图形进行拖动而建立模型，而不是通过元件的数学原理，应用起来十分的方便[50]。该软件一般是应用在液压控制方面，软件中的图标符号基本上都是在实际画图中广泛应用的工程领域标准图符。AMESim内部有丰富的液压模型库，可根据需求来建模。南京理工大学的黄晓华等人应用该软件实现了盾构机的液压系统建模与仿真，国防科技大学周小军等人实现了基于AMESim的液压系统泄漏仿真[51]。仿真图如图5.2所示：

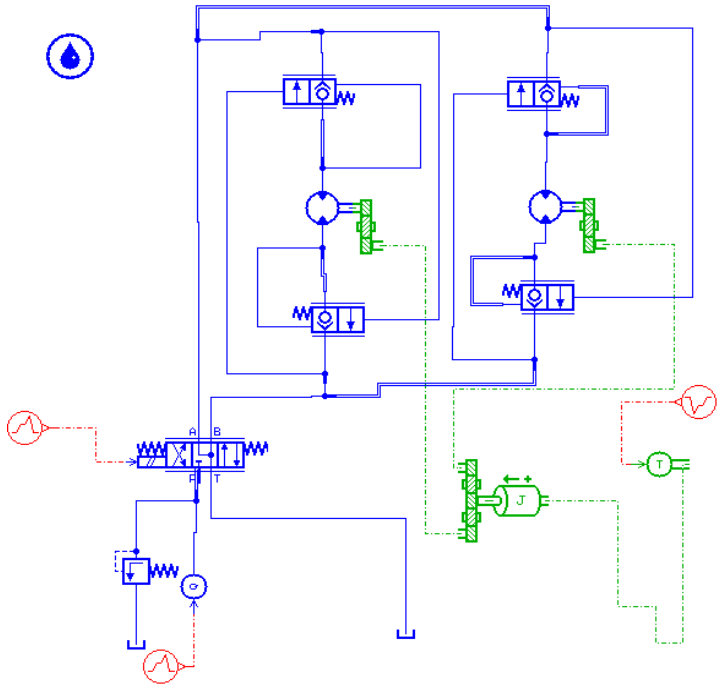

图5.2 周向回转液压回路 AMESim 仿真图

基本的参数主要包括流体的类型、油液温度、油液编号、油液的通用性质等。参数设置如下图 5.3 所示：

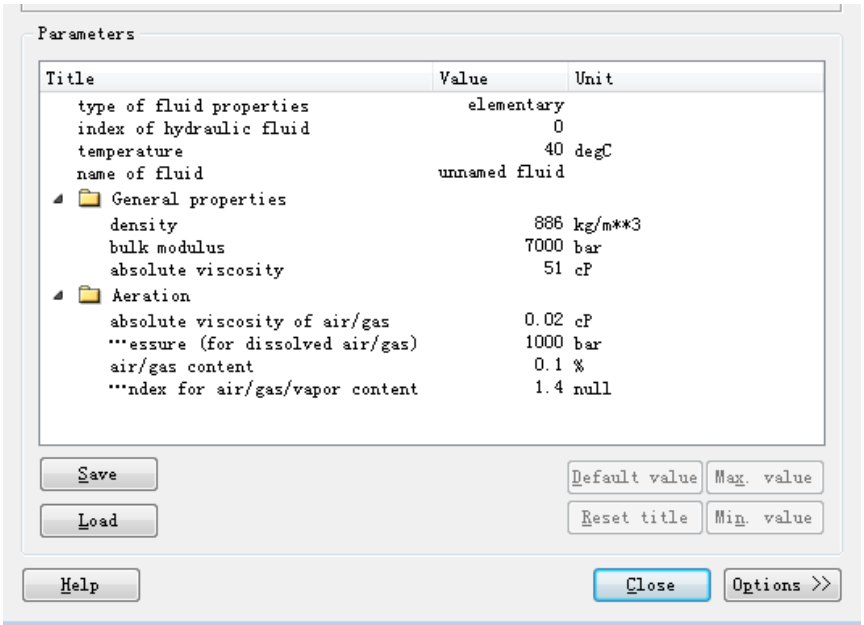

图5.3 AMESim 仿真基本参数图

5.3.2 液压系统油液污染故障仿真

通过对 AMESim 中设置不同的 air/gas content 参数，研究油液若渗入了气体，气

体的含有量大小对液压系统的影响。拼装机在工作时候，不可避免的要接触到空气，油液中的空气含量自然就会增多，为了模拟该故障模式，AMESim 中设置不同的 air/gas content 参数。air/gas content 参数分别设置为 0.1%（正常情况）、10%（轻微故障）、20%（严重故障）。当油液中气体含量不同的时候，马达的相关变量会发生变化，通过马达中变量的分析实现对整个液压系统的故障诊断。本文中主要对马达的第一个端口的流速和压力以及轴扭矩这三个参数变量进行数据采集和观察，通过他们来对整个液压系统进行故障诊断。

（1）正常情况下，马达的参数曲线如图 5.4 所示：

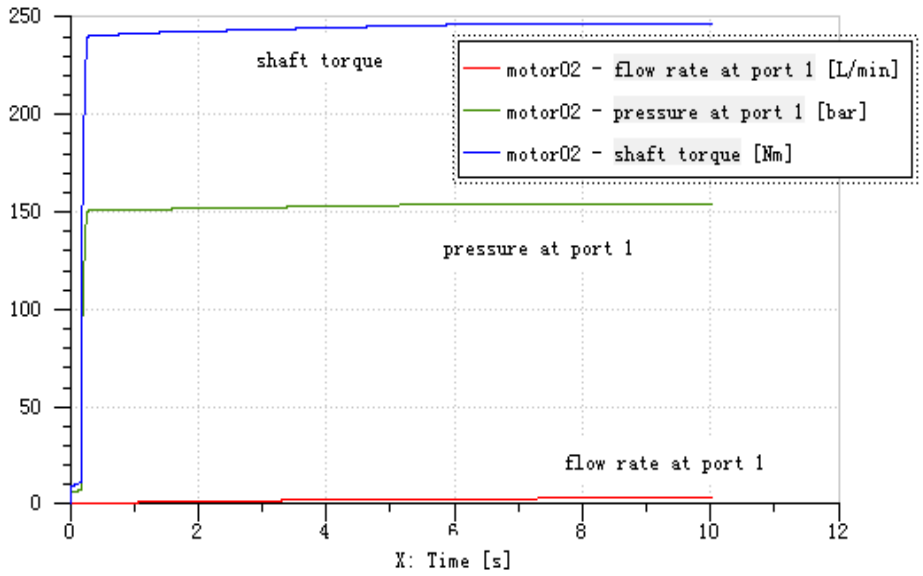

图5.4 油液污染时马达正常情况下的参数曲线图

（2）气体含量为 10%时候，端口处的流速发生了变化，压力和扭矩和正常情况下区别不是很大，马达的参数曲线如图 5.5 所示：

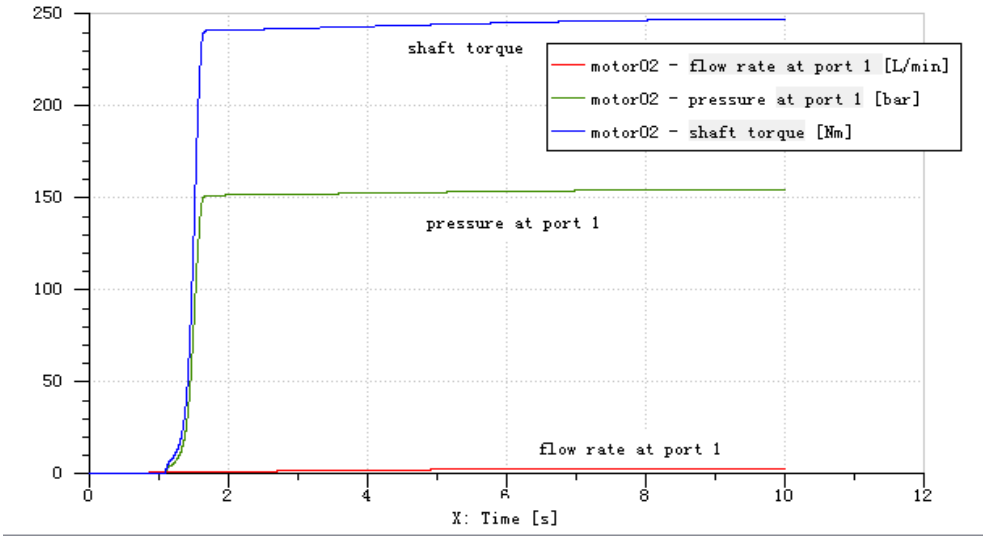

图5.5 油液中气体含量为 10%时马达参数曲线图

（3）当油液中气体含量为 20%时，这时候已经可以看做是严重故障，三个参数的数据都发生了剧烈的变化，马达的参数曲线如图 5.6 所示：

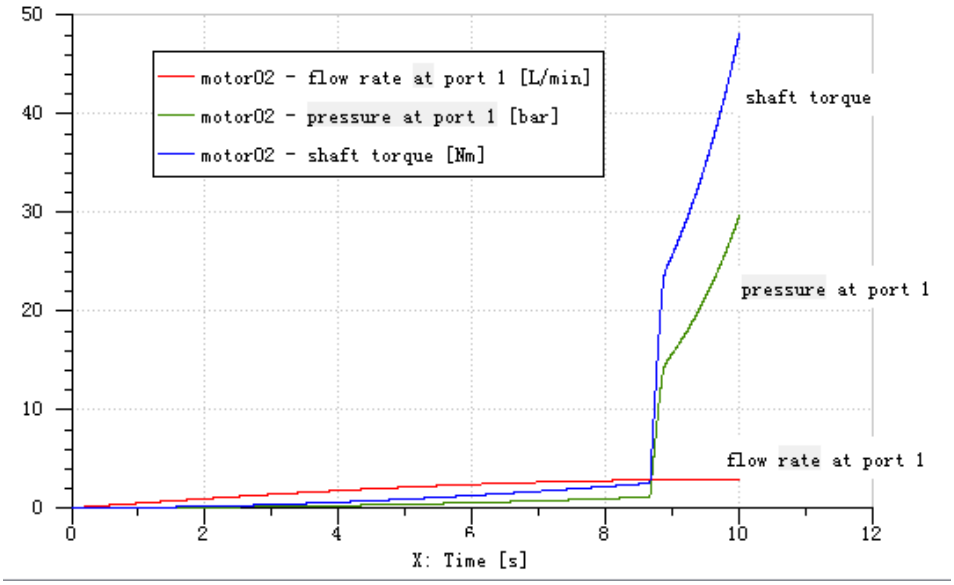

图5.6 油液中气体含量为 20%时马达参数曲线图

5.3.3 液压系统漏油故障仿真

液压油在液压元件容腔里面流动，由于设备设计、工艺以及设备运行时候的压力等问题，会出现油液泄露问题。通过改变 AMEsim 中模型的输入流量来模拟泄漏故障，通过马达压力、转速等因素来判断故障。为了使仿真的效果明显，所

以设置一个比较大的流量值作为正常施工的情况。设定流量参数为 500L/min 作为正常情况，400L/min 为轻微漏油，300L/min 为严重漏油。

（1）当将流量输入设置为 500L/min 时的马达各个参数值曲线如图 5.7 所示：

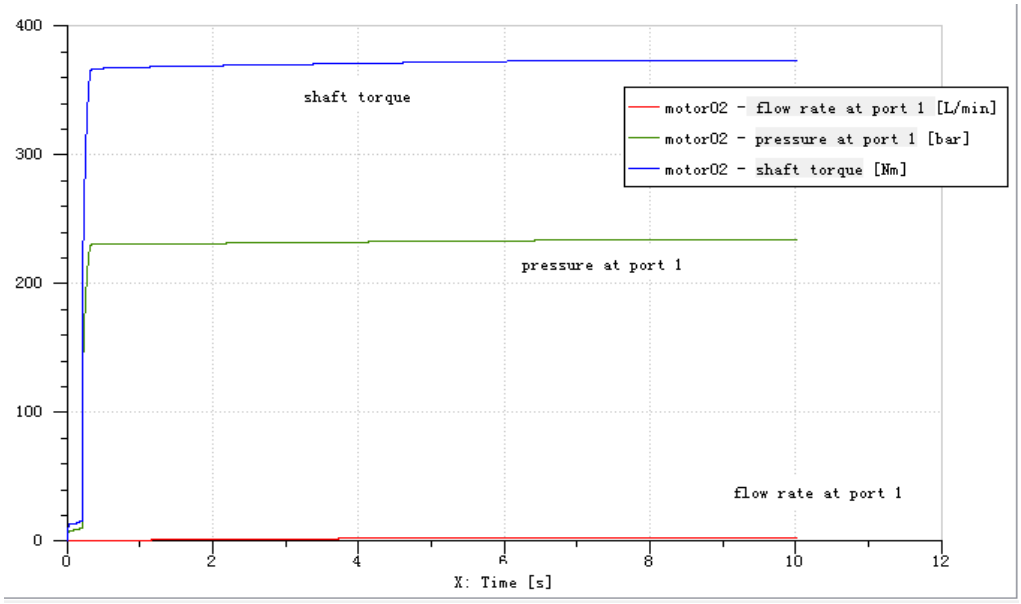

图5.7 输入流量为 500L/min 时马达参数曲线图

（2）当输入设置为 400L/min 时，可以看出旋转扭矩和压力都发生了变化，马达的参数曲线图如 5.8 所示：

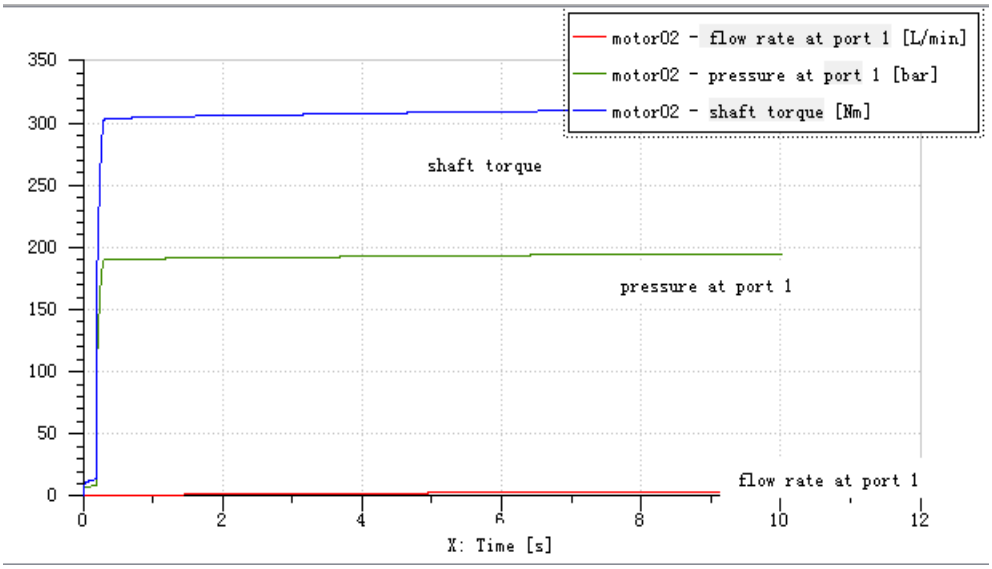

图5.8 输入流量为 400L/min 时马达参数曲线图

（3）当输入为 300L/min 时，已经可以看做是严重漏油。通过观察变量发现压力

和扭矩已经明显不足，马达参数曲线如图 5.9 所示：

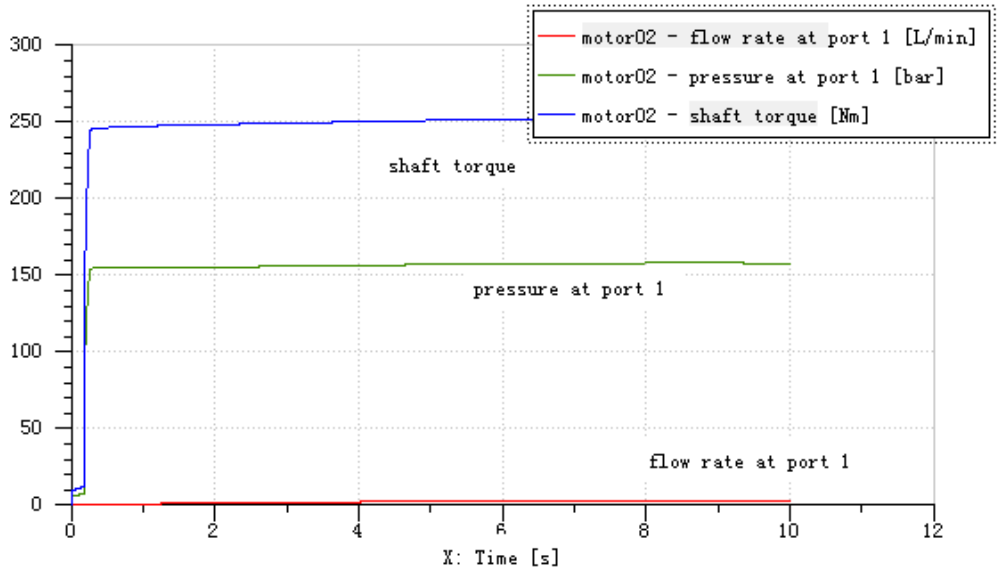

图5.9 输入流量为 300L/min 时马达参数曲线图

5.4 应用基于数据驱动的方法进行故障诊断

5.4.1 油液污染故障仿真分类结果

首先对上文中的仿真图中进行数据的采集，AMESim 软件可以设定采样的频率，并且能够以 txt 格式来对数据进行保存。采用的是随机采样，采集到的数据类型为数值型数据，分别代表马达的扭矩、油液的流速和压力。

首先研究油液中气体含量对液压设备的影响，将故障分为三类正常、轻度故障、严重故障。如第三章在介绍 SVM 的时候提到的，需要考虑核函数问题，在本章中对多项式核函数和高斯核函数进行了对比研究。首先对多项式核函数 $K(x,y)=(<x,y>+1)^p$，再应用高斯核函数 $K(x,y)=e^{-\gamma\|x-y\|^2}$，分别进行实验，分类函数为 $f(x)=\mathrm{sgn}(\sum_{i=1}^{n}a_i y_i K(x,x_i)+b)$。

（1） 核函数为多项式核函数 $K(x,y)=(<x,y>+1)^p$ 时算法分析结果：

表5.1 油液污染时多项式核模型性能

指标	数值	百分比
正确分类样本数	139	92.6667%
错误分类样本数	11	7.33333%
总样本数	150	

表5.2 油液污染时多项式核分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area

正常	0.84	0.03	0.933	0.84	0.884	0.905
轻微故障	0.94	0.08	0.855	0.94	0.895	0.945
严重故障	1	0	1	1	1	1
Weighted Avg	0.927	0.037	0.929	0.927	0.926	0.95

表5.3 油液污染时多项式核函数分类情况

实际类\预测类	正常	轻微故障	严重故障
正常	42	8	0
轻微故障	3	47	0
严重故障	0	0	50

（2）核函数为高斯核函数 $K(x,y)=e^{-\gamma\|x-y\|^2}$ 时算法分析结果：

表5.4 油液污染时高斯核模块性能

指标	数值	百分比
正确分类样本数	143	95.3333%
错误分类样本数	7	4.6667%
总样本数	150	

表5.5 油液污染时高斯分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
正常	0.92	0.03	0.939	0.92	0.929	0.964
轻微故障	0.94	0.04	0.922	0.94	0.931	0.951
严重故障	1	0	1	1	1	1
Weighted Avg	0.953	0.023	0.953	0.953	0.953	0.972

表5.6 油液污染时高斯核函数分类情况

实际类\预测类	正常	轻微故障	严重故障
正常	46	4	0
轻微故障	3	47	0
严重故障	0	0	50

5.4.2 漏油故障数据：

（1）核函数为多项式核函数时算法分析结果：

表5.7 漏油故障时多项式核模型性能

指标	数值	百分比
正确分类样本数	142	95.9459%
错误分类样本数	6	4.0541%
总样本数	148	

表5.8 漏油故障时多项式核分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
正常	0.938	0.03	0.938	0.938	0.938	0.969
轻微故障	0.94	0.031	0.94	0.94	0.94	0.955
严重故障	1	0	1	1	1	1
Weighted Avg	0.959	0.02	0.959	0.959	0.959	0.975

表5.9 漏油故障时多项式核函数分类情况

实际类\预测类	正常	轻微故障	严重故障
正常	45	3	0
轻微故障	3	47	0
严重故障	0	0	50

（2）核函数为高斯核函数时算法分析结果：

表5.10 漏油故障时高斯核模型性能

指标	数值	百分比
正确分类样本数	138	93.8776%
错误分类样本数	9	6.1224%
总样本数	147	

表5.11 漏油故障时高斯核分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
正常	0.915	0.05	0.896	0.915	0.905	0.954
轻微故障	0.9	0.041	0.918	0.9	0.909	0.929
严重故障	1	0	1	1	1	1

Weighted Avg	0.939	0.03	0.939	0.939	0.939	0.961

表5.12 漏油故障时高斯核函数分类情况

实际类\预测类	正常	轻微故障	严重故障
正常	43	4	0
轻微故障	5	45	0
严重故障	0	0	50

5.4.3 三种不同状态的分类结果

在本节中处理的数据集包含有正常状态、轻微油液污染下的状态和漏油状态下的数据，主要是对多分类问题进行验证。

(1) 核函数为多项式核时算法的分析结果：

表5.13 多故障多项式核模型性能

指标	数值	百分比
正确分类样本数	141	94%
错误分类样本数	9	6%
总样本数	150	

表5.14 多故障多项式核分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
正常	0.88	0.03	0.936	0.88	0.907	0.925
油液污染	0.94	0.06	0.887	0.94	0.913	0.955
漏油	1	0	1	1	1	1
Weighted Avg	0.94	0.03	0.941	0.94	0.94	0.96

表5.15 多故障多项式核函数分类情况

实际类\预测类	正常	油液污染	漏油
正常	44	6	0
油液污染	3	47	0
漏油	0	0	50

(2) 核函数为高斯核函数时算法的分析结果：

表5.16 多故障高斯核模型性能

指标	数值	百分比
正确分类样本数	143	95.33333%
错误分类样本数	7	4.6667%
总样本数	150	

表5.17 多故障高斯核分类器性能

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
正常	0.92	0.03	0.939	0.92	0.929	0.945
油液污染	0.94	0.04	0.922	0.94	0.931	0.985
漏油	1	0	1	1	1	1
Weighted Avg	0.953	0.023	0.953	0.953	0.953	0.97

表5.18 多故障高斯核函数分类情况

实际类\预测类	正常	油液污染	漏油
正常	46	4	0
油液污染	3	47	0
漏油	0	0	50

5.5 利用数据挖掘来发现知识

5.5.1 利用关联规则算法来发现知识

在漏油故障和油液污染故障仿真数据集中进行知识发现，首先要对数据进行预处理，将三个数值型变量转化字符变量，分别对变量 shaft torque, pressure, class 规定阈值，若超过或者低于该阈值就将其值赋值为 F，在阈值内就赋值为 T。然后应用算法进行规则发现，得到的结果如下表 5.19：

表5.19 关联规则知识发现

序号	规则	置信度(%)	支持度(%)
1	[shaft torque = T, pressure=T, class = 油液污染]=>[flow rate = F]	92.6	10
2	[flow rate = T, shaft torque = F, class = 漏油]=>[pressure = F]	90.4	9.7
3	[flow rate = T, shaft torque = F, pressure = F]=>[class = 漏油]	90.3	9.5

由于选取的数据量比较小，关联规则的作用表现的不是很明显，故选择网络上的购物篮数据来对关联规则进行验证。当数据量大的时候，可以看出关联规则依旧有效。结果如表 5.20 所示：

表5.20 购物篮 Apriori 算法知识发现结果

序号	规则	置信度(%)	支持度(%)
1	[cannedveg=T beer=T]:167=>[sex=M]:150	89.82	15
2	[cannedveg=T frozenmeal=T]:173=>[sex=M]:150	87.86	15.2
3	[sex=M beer=T]:196=>[cannedveg=T]:150	76.53	15
4	[sex=M cannedveg=T]:199=>[frozenmeal=T]:169	76.38	15.2
5	[sex=M cannedveg=T]:199=>[beer=T]:150	75.38	15
6	[sex=M frozenmeal=T]:209=>[cannedveg=T]:135	72.73	15.2
7	[frozenmeal=T]:302=>[sex=M]:209	69.21	20.9
8	[fish=T]:292=>[homeown=NO]:199	68.15	19.9
9	[beer=T]:293=>[sex=M]:196	66.89	19.6
10	[cnanedveg=T]:303=>[sex=M]:199	65.68	18.1
11	[confectionery=T]:276=>[sex=F]:181	65.58	19.4
12	[fruitveg=T]:299=>[homeown=NO]:194	64.88	18
13	[wine=T]:287=>[sex=F]:180	62.72	17
14	[beer=T]:293=>[cannedveg=T]:170	58.02	17.3
15	[frozenmeal=T]:302=>[cannedveg=T]:173	57.28	17.3

5.5.2 利用聚类算法发现知识

由于在实际中，拼装机液压系统收集到的数据大部分都是数值型的数据，应用聚类分析来进行分组，形成知识规则。通过聚类算法可以将设备的运行状态区分出来。通过该算法可以得到不同状态下的中心点，这些中心点可以作为知识规则存储在知识库中。当新收集到的数据点离哪个中心点较近，就可大致判断管片拼装系统的状态。

针对本文仿真的三种不同的运行状态（正常、油液污染故障、漏油故障），设备的参数之间的数值总会有不同，通过聚类算法，就能够将三种情况下的数据区分开，他们将会聚类成为三种不同的群簇。利用群簇之间的距离，就可以将三种不同的情况分开。由结果可以看出，聚类效果也很好，完全能够将三种状态分开，而且也能够得到中心点数据，这些中心点数据就可以作为知识存储在知识库中。若中心点为 μ ,半径最大值为 D，某个群簇为 c。知识表示形式可以为规则型：

$$\|x_j^{(i)} - \mu_j\|^2 \leq D^2 \Rightarrow x_j^{(i)} \in c$$

聚类算法进行知识发现运行结果如表 5.21 所示：

表5.21 k-means 算法聚类结果

聚类	聚类标号	0	1	2
	样本数	50	50	50
	比例	33.33%	33.33%	33.33%
中心点	flow rate	2.374	2.478	2.778
	pressure	153.462	145.598	155.458
	shaft torque	245.814	236.628	247.864
	class	正常->50(100.0%)	正常->50(100.0%)	正常->0(0.0%)
		油液污染->0(0.0%)	油液污染->0(0.0%)	油液污染->50(100.0%)
		漏油->0(0.0%)	漏油->50(100.0%)	漏油->0(0.0%)

5.6 本章小结

本章分别仿真了正常、油液污染故障和漏油故障。对研究中所涉及到的关键算法 SVM 算法进行了研究然后利用 Apriori 算法和聚类算法 k-means 算法对数据进行分析，得到相应的规则和聚类结果。实验的结果表明，各个算法能够满足要求。

第六章 总结和展望

6.1 研究总结

本文的研究目的是提出一个混合故障诊断模型，并对其中的关键算法进行研究。以盾构机管片拼装液压系统中的周向回路为例，主要完成了对支持向量机算法、关联规则算法和聚类算法的研究。在盾构机的研究中该方面涉及的较少，具有一定的研究价值。本文完成主要研究工作有以下几点：

（1）完成了混合故障诊断模型的总体设计和功能模块设计。

（2）详细阐述了支持向量机算法的原理和流程，讨论了核函数对算法的影响，并对分类问题进行了分析。将该算法应用于盾构机液压系统中，实现对其的故障诊断，研究了算法的性能。

（3）运用关联规则算法和聚类算法在盾构机故障数据中完成了知识发现。对关联规则中的 Apriori 算法和聚类算法中的 k-means 算法的原理和流程进行了阐述。将这两种算法应用在盾构机的故障数据中，进行知识发现。

6.2 研究展望

工业大数据环境下的设备故障诊断研究内容涉及较广，国内外的研究也都是刚刚起步。由于自身理论水平、条件等限制，论文只是在该方向进行了初步的研究，在各个方面还有很大的提升空间。经过对本文系统的思考，发现仍然有以下几点需要进一步的研究：

（1）本文所用的数据量仍然不够大，该方法的处理速度以及效率是否能得到保证仍需要进一步的研究。

（2）本文的盾构机管片拼装液压系统的数据是通过仿真软件仿真得来，因为真实情况中，故障数据比较难收集到，所以数据具有一定的人为因素。

（3）本文中应用的都是一些理论成熟的算法，只是在理论的应用方向有创新，并没有涉及到理论本身的创新。若以后该方向的研究比较深入，可以考虑应用新型的算法来处理问题。

参考文献

[1] Jay Lee, Hung-An Kao, Shanhu Yang. Service innovation and smart analytics gor Industry 4.0 and big data environment[J]. Percedia CTRP, 2014, 16:3-8.

[2] 王国彪，何正嘉，陈雪峰等．机械故障诊断基础研究“何去何从”[J]．机械工程学报，2013，49(1)：63-72.

[3] 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.

[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] 刘其洪，廖弘毅．旋转机械振动故障诊断专家系统[J]．计算机测量与控制，2014，22(12)：3881-3883.

[8] 陈超，李凌均，雷文平等．基于多源信息融合的旋转机械故障诊断专家系统的研究和实现[J]．制造业自动化，2014，36(10)：16-18+22.

[9] 盛博，邓超，熊尧等．基于图论的数控机床故障诊断方法[J]．计算机集成制造系统，2015，06：1559-1570.

[10] 李晗，萧德云．基于数据驱动的故障诊断方法综述[J]．控制与决策，2011，26(1)：1-9+16.

[11] Xuewu Dai, Zhiwei cao. From Model Signal to Knowledge : A data-driven perspective of fault detection and diagnosis[J]. IEEE TRANSACTIONS ON INDUSTRIAL, 2013, 9(4): 1-12.

[12] 马贺贺．基于数据驱动的复杂工业过程故障检测方法研究[D]．上海：华东理工大学，2013.

[13] Ahmad Alzghoul, Magnus Lofstrand. Increasing availability of industrial systems through data stream mining[J]. Computer& Industrial Engineering, 2011, 60(2): 195-205

[14] Andrew Kusiak, Anoop Verma. Analyazing bearing faults in wind turbines: A data-mining approach[J]. Renewable Energy, 2012, 48: 110-116.

[15] Jay Lee, Jun Ni, Dragan Djurdjanovic. Intelligent prognostics tools and e-maintenance[J]. Computer in Industry, 2006, 57(6): 476-489.

[16] 左庆林．盾构机关键设备状态监测与故障诊断研究[D]．石家庄：石家庄铁道大学, 2014.

[17] 赵华，苏东，乔文生．TBM 主变速箱的状态监测与故障诊断[J]．建筑机械化，2003，6: 44-45+43.

[18] 韩超．数据挖掘在盾构机故障诊断中的应用研究[D]．沈阳：沈阳理工大学, 2011.

[19] Larder B, Azzam H, Trammel C et al. Smith Industries HUMS: changing the M from monitoring to management[J]. Proceeding of Aerospace Conference, IEEE, 2000, 6: 449- 455.

[20] 刘强，柴天佑，秦泗钊．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, 25(6): 801-807+813.

[21] Gorka Azkune, Aitor Almeida.Extending Knowledge-driven activity models through data-driven learning techniques[J]. Expert System with Applications, 2015, 8(4): 3115-3128.

[22] 范明．数据挖掘概念与技术[M]．北京：机械工业出版社, 2001.

[23] 方洪鹰．数据挖掘中数据预处理的方法研究[D]．重庆：西南大学, 2009.

[24] 关大伟．数据挖掘中的数据预处理[D]．长春：吉林大学, 2006.

[25] Pearson K. On Lines and Planes of Closetst Fit to Systems of Point in Space[J].Philosophical Magazine , 1901, 2(6): 559-572.

[26] 李玉珍，王宜怀．主成分分析及算法[J]．苏州大学学报(自然科学版), 2005, 21(1): 32-36.

[27] 张荣梅．智能决策支持系统研究开发及应用[M]．北京：冶金工业出版社, 2003.

[28] 曹存根．面向专家的知识获取[M]．北京：科学出版社, 1998.

[29] 丁世飞，齐丙娟，谭红艳．支持向量机理论与算法研究综述[J]．电子科技大学学报，2011, 40(1): 2-10.

[30] Matthias Rychetsky, Stefan Ortmann, Manfred Glesner. Support vector approaches for engine knock detection[C]. International Joint Conference on Neural Networks, 1999: 969-974.

[31] 胡寿松，王源．基于支持向量机的非线性系统故障诊断[J]．控制与决策，2001，16(5): 617-620.

[32] Ganyun Lv, Haozhong Chen, Haibao Zhang. Fault diagnosis of power transformer based on multi-layer SVM classifier [J]. Electric Power Systems Research, 2005, 74(1): 9-15.

[33] Mercer J. Functions of positive and negative type and their connection withtheory of integral equations[J]. Philosophical Transactions of the Royal Societyof London, 1909, 209: 415-446.

[34] 冯新刚．支持向量机核函数选择方法探讨[D]．南昌：江西理工大学, 2012.

[35] 李红英．支持向量分类机的核函数研究[D]．重庆：重庆大学, 2009.

[36] 郭沫．基于多分类支持向量机的核电站故障诊断技术研究[D]．哈尔滨：哈尔滨工程大学, 2012.

[37] 郑建柏，朱永利，张文浩．支持向量机多分类及其在变压器故障诊断中的应用[J]．中国电力教育, 2007, S1: 399-400.

[38] 沈鼎新，王晓．船舶机械故障诊断专家系统[J]．上海海运学院学报, 1988, 9(3): 49-59.

[39] 金亮亮. 基于故障树的航天器故障诊断专家系统研究[D]. 南京: 南京航空航天大学, 2008.

[40] 樊彬彬, 刘谨, 谈理. 盾构机故障诊断专家系统综合推理机的研制[J]. 机械设计与制造, 2007, 3(9): 111-113.

[41] 吴海桥. 现代大型客机故障诊断专家系统的研究与开发[D]. 南京: 南京航空航天大学, 2002.

[42] 李炳燃, 张金哲. 数据挖掘在设备故障诊断专家系统知识获取中的应用[J]. 科技信息, 2011, 20: 232.

[43] 王爱平, 王占凤, 陶嗣干等. 数据挖掘中常用关联规则挖掘算法[J]. 计算机技术与发展, 2010, 20(4): 105-108.

[44] 张昭涛. 数据挖掘聚类算法研究[D]. 成都: 西南交通大学, 2005.

[45] 管会生. 土压平衡盾构机关键参数与力学行为的计算模型研究[D]. 成都: 西南交通大学, 2008.

[46] 孙志超, 黄晓华. 基于 AMEsim 盾构机管片拼装系统的建模与仿真[J]. 机床与液压, 2013, 41(13): 144-146.

[47] 杨琦. 支持向量机在液压系统故障诊断中的应用研究[D]. 大连: 大连海事大学, 2005.

[48] 王盈. 基于双层模糊支持向量机的液压系统故障诊断[D]. 太原: 太原科技大学, 2013.

[49] 贺湘宇. 挖掘机液压系统故障诊断方法研究[D]. 长沙: 中南大学, 2008.

[50] 李亮. 基于 AMESim 的动车组制动系统仿真研究[D]. 成都: 西南交通大学, 2013.

[51] 周小军. 基于 AMESim 液压系统泄漏仿真与故障诊断研究[D]. 长沙: 国防科学技术大学, 2012.

致谢

本论文是在导师的悉心指导下完成的，从论文的选题到论文的撰写，无不渗透着导师的心血。在此论文完成之际，谨此向课题组的所有老师们致以衷心的感谢和崇高的敬意。

光阴似箭，转眼间两年半的研究生生活即将结束。从研一踏进西电的校门开始研究生的学生生活，这两年半的时间里，我感受到了知名学府那种良好的学习氛围、严谨的学术态度，最重要的是在做人和做事等方面得到了锻炼。对那些帮助过，鼓励过，指导过我的人，心中充满了感激。

感谢孔老师，从论文的开题到定稿，孔老师倾注了大量心血，对论文的研究方向，论文框架有很大帮助。在专业上，孔老师有着严谨的治学态度。生活上，孔老师平易近人，态度和蔼，让人如沐春风。感谢课题组中的马老师、殷老师和常老师在科研道路上对我的指导和帮助。

感谢实验室的朱晓灿、章雄等同学，谢谢你们在我科研遇到困难的时候对我的支持和鼓励。感谢 129 宿舍的程养、裴后宣、任军旗，陪我一起度过了一段快乐，忙碌，难忘的时光，有了你们的陪伴，我的生活更加丰富多彩。感谢在西安的同学们，在孤单寂寞的时候谢谢有你们的陪伴，正是因为有了你们，我更爱这个城市。感谢西电，在这里我深深感受到一种精神，百折不挠、越挫越勇的科研和革命热情，我为自己能成为西电人而万分自豪。

我要郑重感谢我的父母，一对平凡普通的农民，含辛茹苦二十多年，培养出我们家的第一位研究生，这需要付出多少心血和精力。感谢父母，你们是我人生最坚强的后盾，你们是我不断前进的源动力，你们是我心灵最温暖的港湾，你们的微笑和幸福是我最大的快乐和安慰。

最后，感谢所有关心、支持和帮助我的人！

作者简介

1. 基本情况

钟福磊，男，山西运城人，1990 年 10 月出生，西安电子科技大学机电工程学院电子机械科学与技术专业 2013 级硕士研究生。

2. 教育背景

2009.08～2013.07 中央民族大学，本科，专业：自动化
2013.08～2016.1 西安电子科技大学，硕士研究生，专业：电子机械科学与技术

3. 攻读硕士学位期间的研究成果

3.1 发表学术论文

[1] Kong Xianguang, Chang Shing I, Yin Lei,Zhong Fulei. Rapid integrated parametric CAE modeling method of Linear Variable Differential Transformer based on a script template[J]. Advances in Engineering Software, 2015, 86: 13-19. (SCI: CK4JW)

[2] 孔宪光，钟福磊，马洪波．工业大数据环境下的混合故障诊断模型研究[C]. 江苏: 全国机械行业可靠性技术学术交流会, 2015.

3.2 参与科研项目及获奖

项目名称：基于工业互联网的离散制造业大数据分析与协同管理研究（教育部基本业务费大数据项目群）2014-2016

在本项目中主要负责面向智能产品和智能装备的故障预测与健康管理:故障检测、故障诊断。以盾构机为实例，研究盾构机的工作原理，利用大数据分析技术对盾构机重要组成部件进行故障诊断和健康管理，对算法进行编程实现。