

# 华中科技大学

## 本科生毕业设计中期检查汇报

题    目： 基于工业大数据的生产设备故障诊断

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

2018 年 4 月

## 一、 课题研究进度安排

表 1 课题研究进度安排表

学期	周次	工作任务
第一学期	18 周——19 周	接受任务，翻译参考文献，完成开题报告，对课题有初步掌握，完成开题答辩
	20 周——21 周	查询资料，完成文献综述等任务
第二学期	1 周——3 周	完成课程设计之余，搜集资料
	5 周——6 周	接受学院检查进度，完成总体方案设计
	7 周——8 周	完成方案 1，基于决策树的模型构建
	9 周——10 周	完成方案 2，基于支持向量机的模型
	第 11 周	撰写毕业论文，完善资料
	第 12 周	完善论文，并且进行论文查重
	13 周——15 周	论文答辩，并且评定成绩

## 二、毕业设计内容概述

### 1. 预期达到的目标

- 1.1 获取足量数据、实现基于数据驱动的故障模型的建立
- 1.2 能基于故障时的异常数据完成对故障的推理与诊断
- 1.3 能够建立故障数据数据库，不断丰富故障模型
- 1.4 人机交互接口，提供生产人员与故障模型的交互界面

### 2. 关键内容

#### 2.1 数据挖掘

对采集到的数据进行清理，挖掘，形成有价值的知识，赋予相对应的故障信息，使得最后形成可以被理解的相关信息，以此为基础构建故障模型；

#### 2.2 故障数据的处理

一个设备具有很多的参数，如风机，它的参数可能包括电机电流、电机线圈温度、轴承温度、振动值、进出口介质温度和流量等，这些参数间是有复杂的关联关系的。当我们构建故障模型的过程中，必须要通过计算参数间的关联度这种手段剔除一些对设备运行状态影响不大的测点，从而提高整体的诊断精度水平。

#### 2.3 设备运行数据的获取

工业大数据需要海量的生产设备历史数据和实时运行数据，这些都需要通过一定的数据采集手段才能得到，这也是本课题的一个重要问题，即如何获取足量的数据来训练模型，使其达到理想的性能与精度。

### 三、毕业设计已完成部分

#### 1. 数据获取

目前主要的数据获取手段是通过互联网上的共享数据集，当前已经获得的数据集合有两个：

一个是来自罗马的一家通信科学研究所：[Semeion Research Center of Sciences of Communication](#)

数据特性如下：

<b>Data Set Characteristics:</b>	Multivariate	<b>Number of Instances:</b>	1941	<b>Area:</b>	Physical
<b>Attribute Characteristics:</b>	Integer, Real	<b>Number of Attributes:</b>	27	<b>Date Donated</b>	2010-10-26
<b>Associated Tasks:</b>	Classification	<b>Missing Values?</b>	N/A	<b>Number of Web Hits:</b>	55309

第二个数据集来自 Github 上一个 Fault\_Diagnosis 项目的自带的风力涡轮内部齿轮箱数据集。该 Github 项目地址为：[Gearboxdata/Gear-Box-Fault-Diagnosis-Data-Set](#)

该数据集内的数据分为两类，即正常运行数据和故障状态下的数据。每一类数据下又按照 0-90HZ，每 10HZ 一个层次分为 10 种运行状态。合共 20 个文件，一共 2021119 条记录，每条记录包括频率在内共 5 个属性。

BrokenTooth Data.zip	Add files via upload	20 days ago
Healthy Data.zip	Add files via upload	20 days ago
README.md	Create README.md	20 days ago
<div> <div>README.md</div> <div> <h2>Gear-Box-Fault-Diagnosis-Data-Set</h2> </div> </div>		

## 2. 模型构建

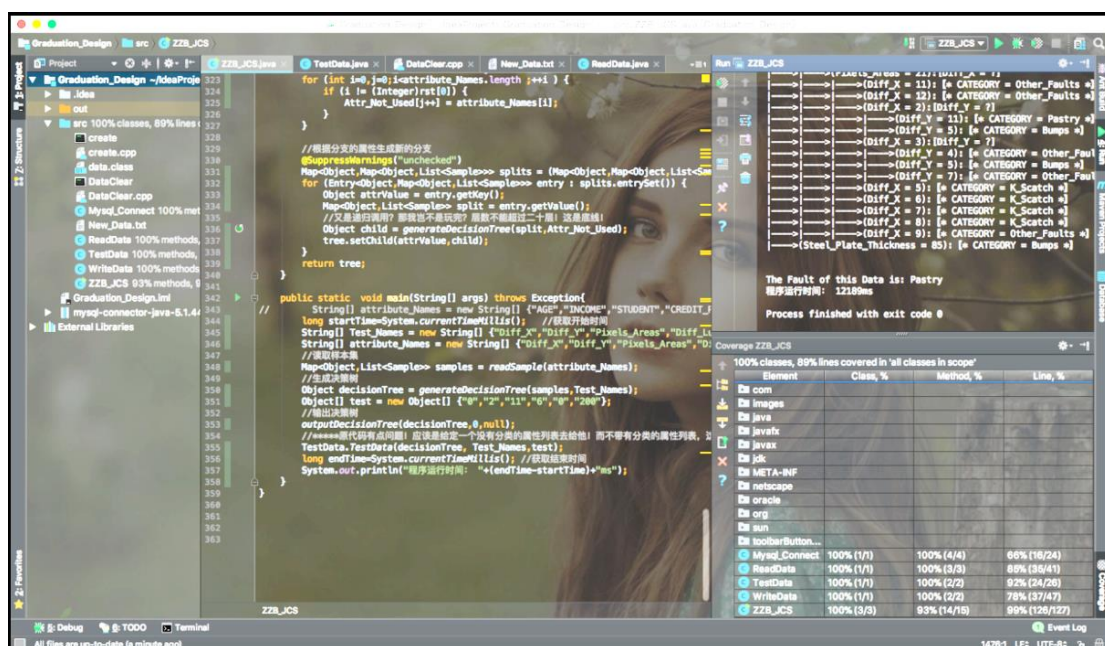
采用决策树算法，在对第一个数据集进行极大地精简之后，终于成功运行出来了一个效率较高，准确度也有保障的故障树模型。

此模型基于 ID3 算法搭建，数据存储于 Mysql 数据库中，整体采用 Java 编写代码。

```
public static void main(String[] args) throws Exception{
    long startTime=System.currentTimeMillis();    //获取开始时间
    String[] Test_Names = new String[]
{"Diff_X","Diff_Y","Pixels_Areas","Diff_Luminosity","TypeOfSteel","Steel_Plate_Thickness"};
    String[] attribute_Names = new String[]
{"Diff_X","Diff_Y","Pixels_Areas","Diff_Luminosity","TypeOfSteel","Steel_Plate_Thickness","Fault
"};

    Map<Object,List<Sample>> samples = readSample(attribute_Names);
    Object decisionTree = generateDecisionTree(samples,Test_Names);
    Object[] test = new Object[] { "0","2","11","6","0","200" };
    outputDecisionTree(decisionTree,0,null);
    TestData.TestData(decisionTree, Test_Names,test);
    long endTime=System.currentTimeMillis(); //获取结束时间
    System.out.println("程序运行时间：  "+(endTime-startTime)+"ms");
}
```

这是整个决策树的主函数，其他调用类、函数定义、数据类型定义合共 600+行，最终运行结果如图。

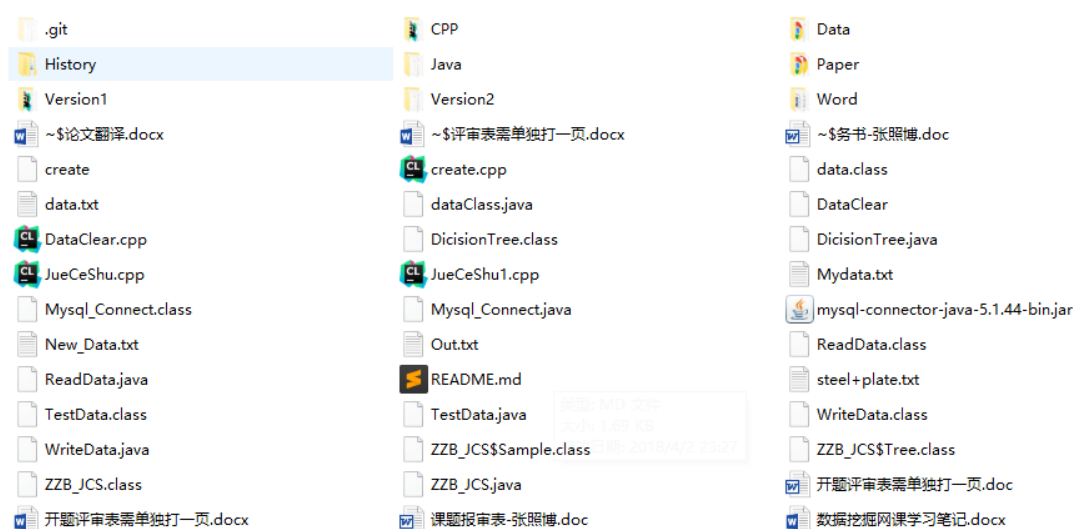


### 3. 方案 2 与人机交互界面

这两个都还在编写过程中，没有实物可以展示。

### 4. 整体完成情况

可以说是完成了一半的工作了，主要就是文献查找、内容理解、编码设计、数据收集、数据处理几个方面都已经涉及了，虽然还有一些工作需要完成，但是对比进度安排表，还是可以发现目前与进度持平的。下一阶段的安排见第四节的后期安排。



## 四、后期时间安排

1. 下周尽量将方案 2 写出来并且测试完毕。
2. 对第二个数据集进行整理并且存储到数据库中，构建新的故障树模型。不过可能计算机资源会消耗过度，所以决策树可能没法实现大规模的故障树建模，只能用一部分作为训练集建模，另外的作为测试集进行模型的准确度测试。
3. 最好是线下获取到一份设备实时运行的数据，这样会比较有实际意义。
4. 如果时间充足，可以尝试在 Hadoop 上对第二个数据集进行分布式 SVM 测试。即将模型写入不同的设备间，分布式分发测试集数据，最后每台机器得到测试结果，并且进行汇总,从而实现一个简单地大数据内容。

## 五、近期参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Zhang, Liangwei. Big Data Analytics for Fault Detection and its Application in Maintenance, 2016
- [5] Jay Lee, Hung-An Kao, Shanhu Yang. Service innovation and smart analytics for Industry 4.0 and big data environment[J]. Percedia CTRP, 2014, 16:3-8.
- [6] 邳文君, 宫秀军. 基于 Hadoop 架构的数据驱动的 SVM 并行增量学习算法[J]. 计算机应用, 2016, 36(11): 3044-3049.
- [7] 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(06): 44-45+43.
- [8] 徐牧. 基于 SVM 的变压器故障诊断研究[D]. 安徽理工大学, 2017

- [9] 罗雨滋,付兴宏. 数据挖掘 ID3 决策树分类算法及其改进算法[J]. 计算机系统应用, 2013, 22(10):136-138+187.
- [10] 张媛. 采用数据挖掘技术中 ID3 决策树算法分析学生成绩[J]. 科技信息, 2009(06):537.
- [11] 张睿. ID3 决策树算法分析与改进[D]. 兰州大学, 2010.
- [12] 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.
- [13] 朱霄珣. 基于支持向量机的旋转机械故障诊断与预测方法研究[D]. 华北电力大学, 2013.
- [14] 易辉. 基于支持向量机的故障诊断及应用研究[D]. 南京航空航天大学, 2011.
- [15] 王振华, 杜宇波. 基于 ESMD 和 SVM 的滚动轴承故障诊断[J]. 现代制造技术与装备, 2018(01):122+124.
- [16] Yang Li, Yan Qiang Li, Zhi Xue Wang. Fault Diagnosis of Automobile ECUs with Data Mining Technologies[J]. Applied Mechanics and Materials, 2011, 1069(40).
- [17] Xiao Rong Cheng, Qiong Wang. An Improved ID3 Algorithm for Power Equipment in Green Power Engineering[J]. Applied Mechanics and Materials, 2013, 2488(340).
- [19] Huan Huang, Natalie Baddour, Ming Liang. Bearing fault diagnosis under unknown time-varying rotational speed conditions via multiple time-frequency curve extraction[J]. Journal of Sound and Vibration, 2017
- [20] Guo Ping Li, Qing Wei Zhang, Ma Xiao. Fault Diagnosis Research of Hydraulic Excavator Based on Fault Tree and Fuzzy Neural Network[J]. Applied Mechanics and Materials, 2013, 2308(303).