

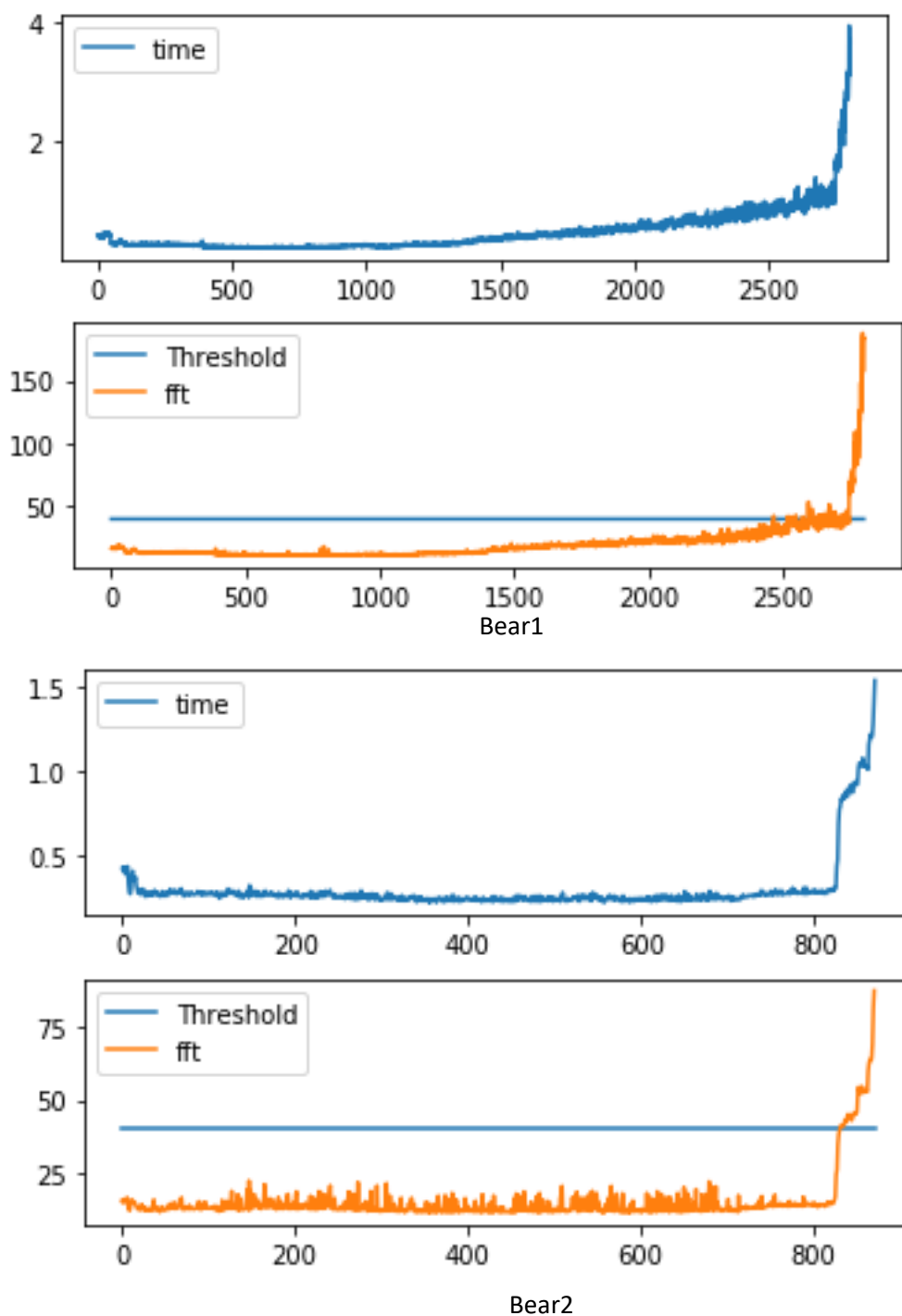
轴承的剩余寿命预测分析报告

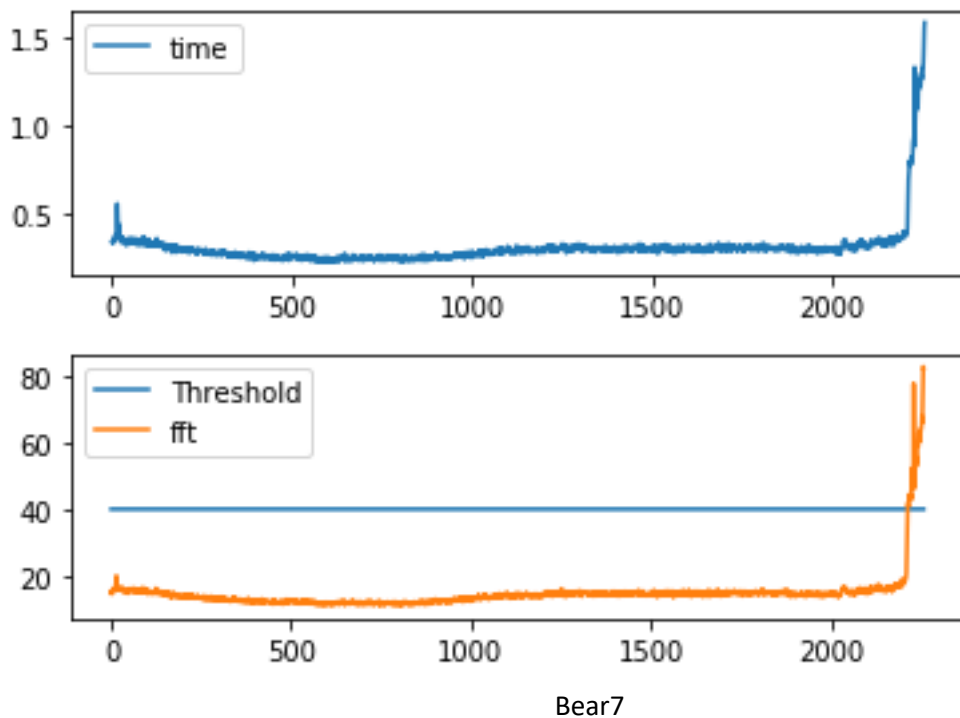
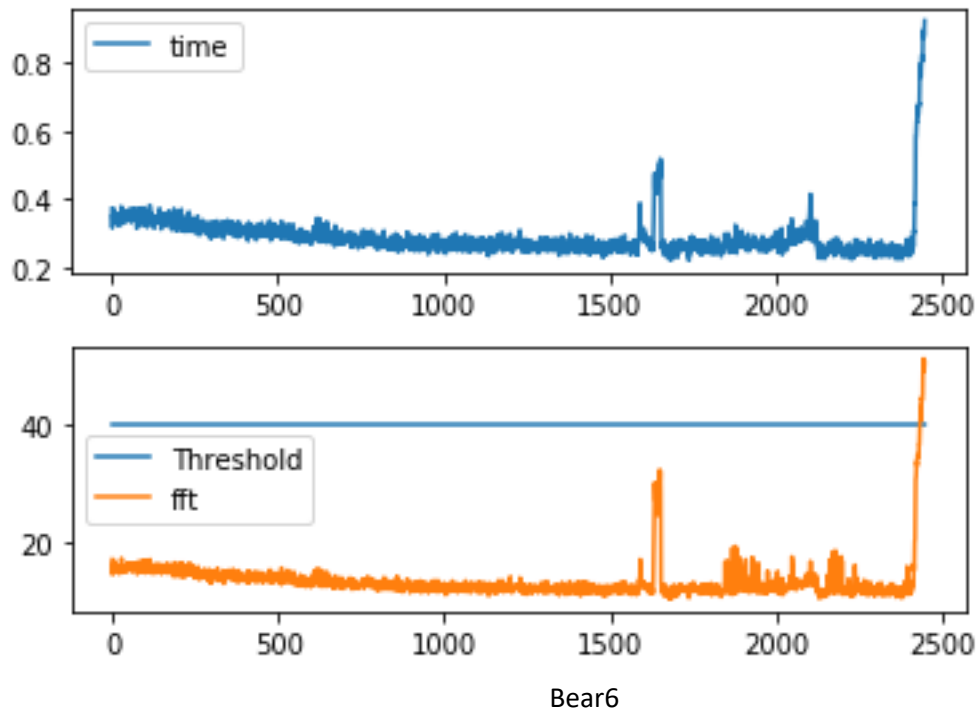
查阅相关论文资料，方案的整体思路如下：

利用现有的两个方向的加速数据，来预测未来时间段加速度的数据，再根据现有的数据设定一个固定的阈值，连续时间段大于该阈值该时刻即判定为该轴承失效。

一、数据的预处理

根据所给的数据与特征分析，每个时刻采集 2560 个数据点，所以先做各个时刻加速度的平均值分析，由于轴承的对称性，所以只对一个方向作分析既可以表示垂直方向的特征。各个轴承的可视化分析特征图如下：





以上即为 4 个轴承的数据分析，从时域和频域对比可以看出，频域的特征区分较为明显，更能体现出细节信息，所以选择频域特征作为训练特征。

根据轴承 1、2、6、7 各个轴承的频谱特征图走向，设定阈值 40 较为合适。

方案一：LSTM 时序大法

根据题目描述和数据分析，这是一个时序问题，利用时序算法四个轴承建立四个时序模型，再针对三个测试轴承，分别利用四个模型预测在做加权得到最后的预测结果，再根据之前设定的阈值，大于阈值即为轴承后面时间段即为失效时刻。但是考虑到各个轴承都是时序的数据，需要建立四个时序模型。但是由于数据较少，四个轴承的数据建

立四个模型，存在很大的过拟合风险，根据之前的机器学习相关的经验和由于时间关系先舍弃此方案。

方案二：时序转回归大法

1、特征工程

把整个时序问题看成是一个监督学习的回归问题，利用前 **t-1** 时刻的值作为该样本的特征，**t** 时刻的值作为该样本的 **label**，如下图：

Figure 1: A screenshot of a Jupyter Notebook showing a DataFrame named 'x_train'. The DataFrame has 300 rows (Index 0 to 299) and 17 columns. The columns are labeled 'var1(t-300)', 'var1(t-299)', 'var1(t-298)', 'var1(t-297)', 'var1(t-296)', 'var1(t-295)', 'var1(t-294)', 'var1(t-293)', 'var1(t-292)', 'var1(t-291)', 'var1(t-290)', 'var1(t-289)', 'var1(t-288)', 'var1(t-287)', 'var1(t-286)', 'var1(t-285)', and 'var1(t-284)'. The values in the columns are numerical, representing time series data points.

特征图

Figure 2: A screenshot of a Jupyter Notebook showing a Series named 'y_label'. The Series has 300 rows (Index 0 to 299) and one column labeled 'var1(t)'. The values in the column are numerical, representing the target values for the time series data.

Label 图

2、模型的选择

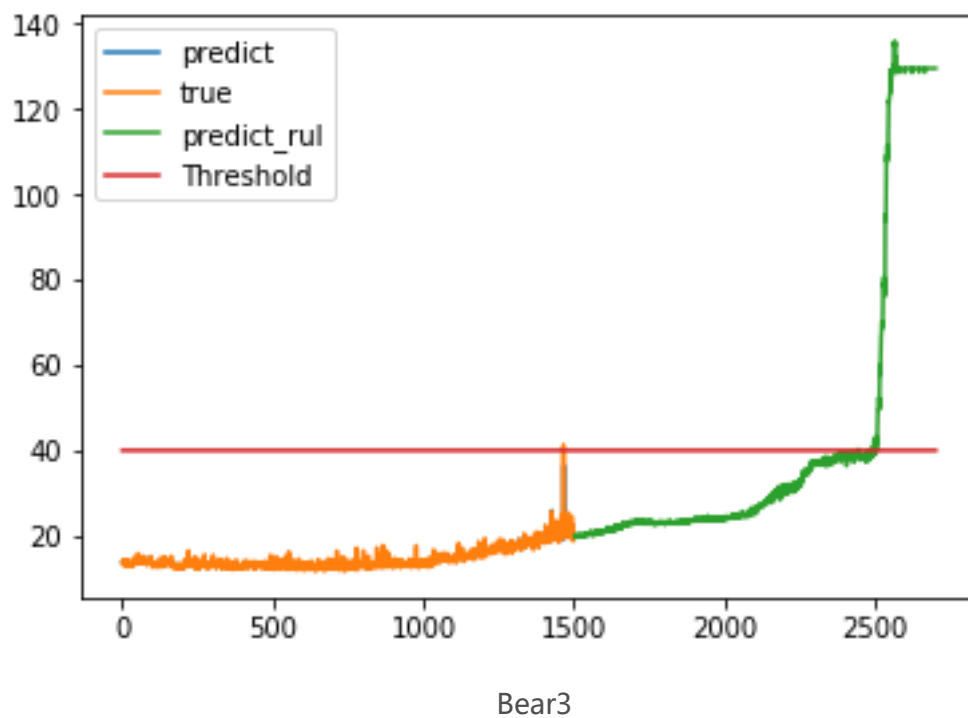
根据特征工程的特征，应该选择为回归的算法模型，结合各种回归算法的特点。最后选择以 CART 回归树为基树的 LightGBM 模型。LightGBM 的主要优势有：

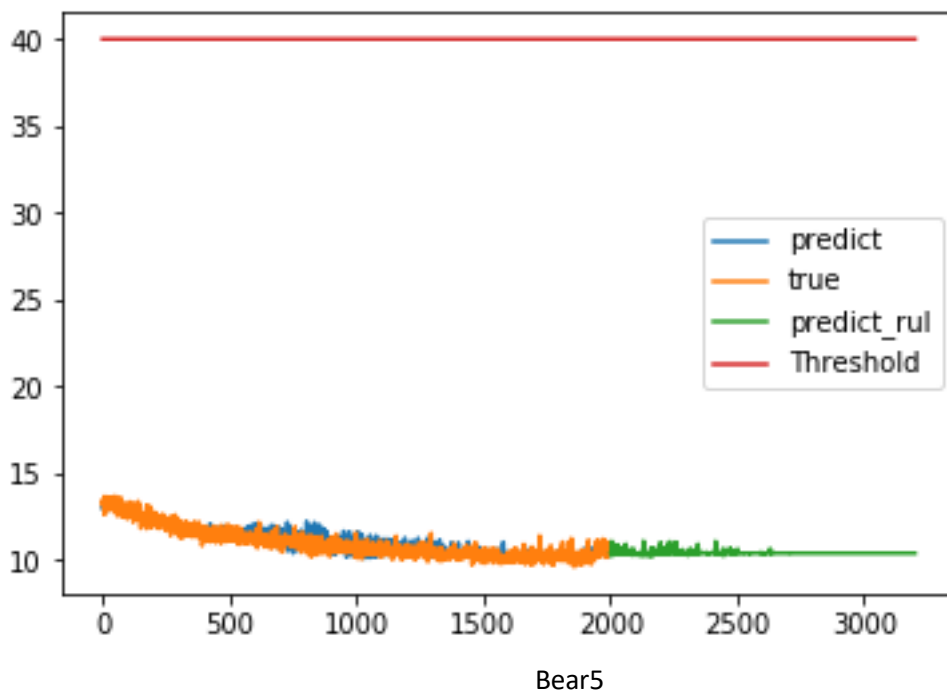
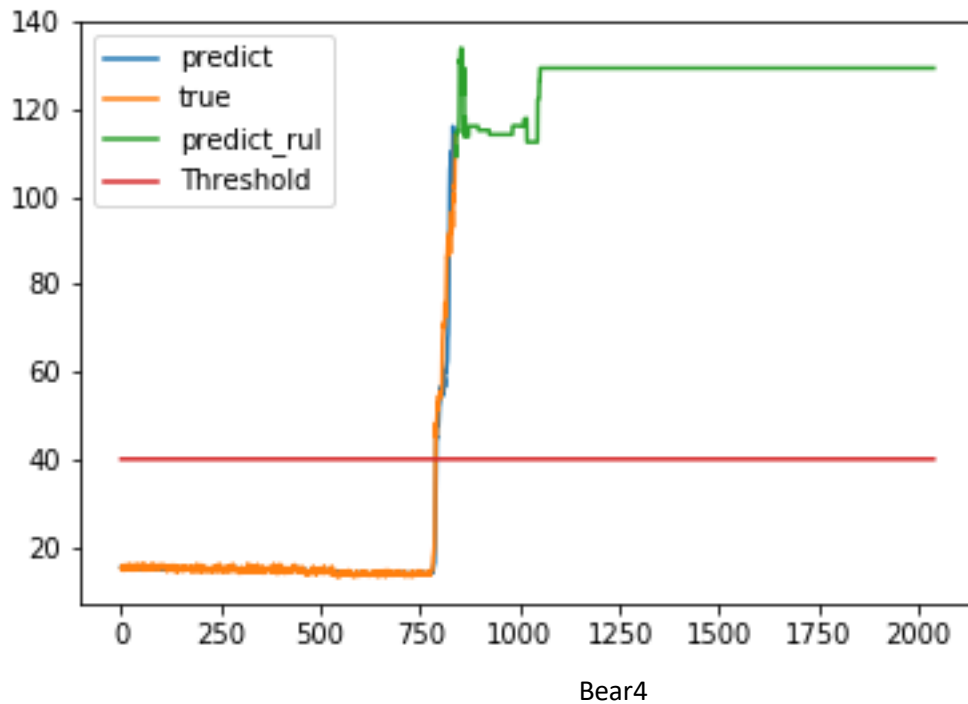
一、更快的训练速度和更高效的效率。

二、更低的内存占用。

三、更高的准确率。

根据 LightGBM 对 3、4、5 轴承的预测效果图如下：





利用均方差来评估模型的效果，选择均方差的原因是因为均方差可以综合评估数据分布的偏差和方差的大小。理想条件下均方差为 0，方差和偏差都为 0，模型完美拟合。主要利用均方差的绝对值来评估模型的效果，均方差绝对值越小，模型的拟合效果越好。

最后模型的训练得到的均方差为：**-0.13052219347108704**。

结论：

轴承 3 剩余寿命还有 10000S，轴承 4 已经失效，轴承 5 预测失败。

轴承 5 预测失败的原因分析：

主要因为轴承 5 的加速度和训练集轴承的加速度对比非常异常，刚开始前期加速度

变化率十分剧烈，后面慢慢的波动减小，根据查阅轴承的特性，有可能是刚刚装配的处于磨合状态的新轴承数据。然而训练数据有限没有这中数据分布的轴承数据，训练数据和预测数据的分布相差太大，所以导致轴承 5 预测失败。

后期可以改进的方向：

- 一、提取的特征太少，可以加入时域的特征和频域的其他特征。
- 二、结合时序算法和回归算法综合预测寿命。
- 三、考虑深度学习模型可能可以得到很好的效果。