

# 飞机发动机剩余使用寿命预测

## 摘要

针对飞机发动机剩余使用寿命预测问题，本文基于传感器监测数据，采用机器学习方法建立了预测模型。对于问题一，分析

关键词：剩余使用寿命；机器学习；集成学习；预测性维护

---

## 一、问题重述

### 1.1 研究背景

飞机发动机作为航空器的核心动力装置，其运行可靠性直接关系到飞行安全。在长期服役过程中，发动机关键部件如高压压气机叶片等，随着运行时间的增加，会逐渐发生性能退化。传统维护策略多采用定期检修方式，存在维护过度或不足的问题。近年来，随着传感器技术和数据分析方法的发展，基于状态的预测性维护逐渐成为主流。

### 1.2 问题描述

本题提供了两组发动机运行数据：

问题一：100台发动机在单一工况下因HPC性能退化导致失效的传感器记录，要求建立模型预测测试集中各发动机的剩余循环寿命。

问题二：100台发动机因HPC或风扇故障失效的传感器数据，需考虑安全惩罚函数，建立预测模型。

### 1.3 评价指标

均方根误差：

$$RMSE = \sqrt{1/N * \sum (y_i - \hat{y}_i)^2}$$

安全惩罚函数：  $S = \sum s_i$

其中： - 当  $d_i < 0$  时，  $s_i = \exp(-d_i/13) - 1$  - 当  $d_i \geq 0$  时，  $s_i = \exp(d_i/10) - 1$

这里  $d_i = \hat{y}_i - y_i$  为预测误差。该函数对预测值大于真实值（即预测过晚）的情况惩罚更重，体现了安全性优先的原则。

---

## 二、数据分析

### 2.1 数据结构

训练数据包含以下字段： - Unit\_ID：发动机编号 - Time\_Cycle：累计运行周期 - Setting\_1~3：操作设置参数 - s1~s21：21个传感器测量值

训练集为完整的运行至失效数据，每台发动机从某一周期开始记录直至失效。测试集则在失效前某一时刻截断，需要预测剩余寿命。

### 2.2 数据特点

通过对训练数据的统计分析发现： 1. 不同发动机的总寿命差异较大，问题一中寿命范围为128-362周期 2. 部分传感器数值基本恒定，不具有区分性 3. 有效传感器的数值随运行周期呈现规律性变化趋势

### 2.3 传感器筛选

计算各传感器与RUL的相关系数，筛选出相关性较强的传感器用于建模。以问题一为例，选取的主要传感器包括：s2、s3、s10、s11、s12、s13、s14、s15、s16、s17、s18、s19、s20、s21。

---

### 三、模型构建

#### 3.1 RUL计算

对于训练数据中的每条记录，计算其RUL值： $RUL = T_{max} - t$

其中  $T_{max}$  为该发动机的最大运行周期（失效时刻）， $t$  为当前周期。

考虑到RUL值过大时预测意义不大，设置上限阈值（问题一为125，问题二为150）进行截断。

#### 3.2 特征工程

针对筛选出的传感器，提取以下特征：

1. 移动平均：计算5周期和10周期的移动平均值，平滑噪声干扰
2. 变化量：一阶差分，反映传感器值的变化速率
3. 累计偏差：当前值与初始值的差值，表征退化程度
4. 周期对数：对运行周期取对数，增加非线性特征

#### 3.3 模型选择

采用集成学习方法，组合多个基学习器：

1. 随机森林：对高维特征有较好的处理能力，不易过拟合
2. 梯度提升树：通过迭代优化逐步降低预测误差
3. XGBoost：在梯度提升基础上进行了多项优化
4. LightGBM：采用直方图算法，训练效率更高

最终预测采用加权平均： $\hat{y} = \sum(w_k * f_k(x))$

各模型权重根据验证集表现确定。

#### 3.4 安全性调整

考虑到安全惩罚函数对预测过晚的严厉惩罚，在最终预测结果上进行保守修正，适当下调预测值，以降低预测过晚带来的安全

---

### 四、结果分析

#### 4.1 问题一结果

指标	数值
验证RMSE	1.53
预测最小值	5
预测最大值	123
预测均值	74.9

#### 4.2 问题二结果

指标	数值
验证RMSE	1.55
预测最小值	3
预测最大值	147
预测均值	80.1

---

## 五、总结

本文针对发动机剩余寿命预测问题，通过数据分析、特征工程和模型构建，建立了较为有效的预测模型。主要工作包括：

1. 分析了传感器数据特点，筛选出与退化相关的有效特征
2. 采用多种时序特征提取方法，丰富了模型输入
3. 使用集成学习方法，综合多个模型的预测能力
4. 针对安全惩罚函数特点，进行了保守性调整

后续可考虑引入深度学习方法（如LSTM）更好地利用时序信息，进一步提升预测精度。

---

## 参考文献

- [1] Saxena A, Goebel K. Turbofan engine degradation simulation data set[R]. NASA Ames Prognostics Data Repository, 2008.
- [2] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [3] Chen T, Guestrin C. XGBoost: A scalable tree boosting system[C]. KDD, 2016: 785-794.
- [4] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012.
- 

## 附录：源代码

见 rul\_predict.py 文件。