1. 整理数据(2503)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 民勤 |  |  | train | test |
| 70022 | 发电机后轴承温度超限 | 发电机 | 16 | 4 |
| 30071 | 变流器通讯心跳超时 | 变流器 | 26 | 7 |
| 110101 | 塔底柜子温度超限 | 机舱 | 30 | 8 |
| 80072 | 齿轮箱润滑油泵马达保护 | 齿轮箱 | 22 | 6 |
| 50035 | 液压系统主压力过低 | 液压系统 | 18 | 5 |
| 60300 | 偏航软启动器故障 | 偏航系统 | 27 | 7 |
| 3011456 | 变桨系统轴123电容柜欠电压 | 变桨系统 | 10 | 2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 红骥 |  |  |  |  |
| 30071 | 变频器RS232通讯中断 | 变流器 | 46 | 11 |
| 80001 | 齿轮箱低速轴温度超限 | 齿轮箱 | 42 | 10 |
| 800020 | 变桨系统紧停动作 | 变桨系统 | 42 | 10 |
| 10002 | 机舱振动瞬时值超限 | 机舱 | 24 | 5 |
| 70021 | 发电机前轴承温度超限 | 发电机 | 5 | 2 |
| 60006 | 偏航齿轮润滑 | 偏航系统 | 14 | 3 |
| 50035 | 液压系统主压力过低 | 液压系统 | 2 | 1 |

2. 数据读取及处理

1. 读取原数据并存储；
2. 替换上下限；
3. 数据分割：环境，设定，状态
4. 数据标准化：
5. 环境（风速+风向），风向区分方向，要加一个方向符号；风速区分大小，用最大最小即可；因此用minmax\_stan4
6. 设定，功率因数要加方向符号，其他普通即可，因此用minmax\_stan4
7. 状态，分为两部分V1+V2，V1有方向的正负号，用minmax\_stan4，V2正负号无方向，用minmax\_stan5
8. 数据白化：
9. 计算mu, std, u, s, v
10. 白化v
11. 合并数据

3. 特征提取

1）维数提取（vae）

2）时间压缩（attention）（lstm-vae）

4．聚类（前3种论文给出来的方法）

1）MiniBatch Kmeans：得到聚类中心C和离聚类最近的nn个样本点索引clsnn

1. 拟合聚类：mbk.fit(Min)
2. 得到聚类中心：C = mbk.cluster\_centers\_
3. 计算样本点与各聚类中心的距离：mcdist = mc\_distance(Min, C, distype='euclideane')
4. 得到离各聚类中心最近的nn个样本点的索引（从近到远排）：clsnn = np.argsort(mcdist, axis=0)[0: nn]

2）Agglomerative Clustering

算法介绍：https://blog.csdn.net/lanchunhui/article/details/50877161

3）BIRCH

4）DBSCN

5）Gaussian Mixture

1. 处理数据

80001， （50, 345, 384）50p+200n，