**Nature Energy数据集中的F1-F9，如果要选择5个互相独立性强的特征出来，应该选择哪几个？这两天你做一下，讲一下除掉的几个特征的理由，除掉的特征按照重要顺序排个序。**

我会选择F2, F5, F3, F6, F9.

|  |  |  |
| --- | --- | --- |
| 特征 | 含义 | 物理意义 |
| F2 | 第100圈和第10圈放电容量随电压差值的方差  **the log variance of ΔQ100-10 (V)**  ΔQ100-10(V):”Difference of the discharge capacity curves as a function of voltage between the 100th and 10th cycles” | 电池反应过程中过电势的早期变化 |
| F5 | 第二圈放电容量  **Discharge capacity, cycle 2** | 电池的初始状态（容量） |
| F9 | 第100圈和第2圈间的内阻增量  **Internal resistance, difference**  **between cycle 100 and cycle 2** | 电池的初始状态及早期变化（电解液的浸润状态，活性物质与集流体的接触） |
| F6 | 前五圈的平均充电时间  **Average charge time, first 5**  **cycles** | 电池的使用工况 |
| F3 | 放电容量曲线第2圈到第100圈线性拟合的斜率  **Slope of the linear fit to the**  **capacity fade curve, cycles 2**  **to 100** | 电池的容量衰退速率 |

按重要性排序，依次除掉F1, F4, F8, F7。

|  |  |  |
| --- | --- | --- |
| 特征 | 含义 | 物理意义 |
| F1 | 第100圈和第10圈放电容量随电压差值的最小值 | 电池反应过程中过电势的早期变化，与F2强相关（0.996） |
|  |  |  |
| F4 | 放电容量曲线第2圈到第100圈线性拟合的截距 | 电池的初始状态（容量），可由F3和F5综合得出（F4和F5：0.938） |
| F8 | 第2圈到第100圈的内阻最小值 | 最小内阻；最小内阻有的数据点为0，且与F6应该有较强的关联性(0.291) |
| F7 | 第2圈到第100圈的温度积分 | 电池的使用工况（温度） |

1. 除掉F1和F4是因为这两个特征分别和F2，F5有重复。因为电池的过电势能综合体现电池的衰退情况，所以F1是最重要的。F1和F2都是取自电池的容量-电压曲线，只是索取的统计值不一样，反映的都是和电池过电势有关的物理现象。电池的长存储时间会使电池发生自放电，石墨阳极表面会形成钝化膜，**消耗活性锂**，所以会出现容量下降的情况。而上述的日历老化现象会对电池的循环性能有一定影响。电池的初始容量是衡量电池日历老化程度最直观的指标。F5体现的是电池循环的初始容量。F4对应的是放电容量曲线第2圈到第100圈线性拟合的截距。因为早期放电容量曲线线性拟合的斜率（F3）很小，其所体现的截距和电池的初始容量十分接近;故F4基本可由F5与F3综合得出。
2. 除掉F8，是因为其对应的是电池早期前100圈循环的最小内阻。其对应的只是一个最小值，而且该最小值出现的位置是不确定的，且不是电池循环最开始的值，具有很大的不确定性。而且该特征与F9相比，不能说明电池经历的100圈的循环，其内阻是增大了还是减小了？增大了多少？减少了多少，这些都是未知的。所以，F8是未选特征中重要性较低的，且与F6有较大的关联性。
3. 除掉F7，是因为与所测量的电池温度有关的因素主要就是：充放电倍率，以及恒温箱的散热情况。充放电倍率这个可以通过F6有所体现。该数据集的有72种充电方式，放电倍率统一为4C，可以看作电池的充放电循环都是在高倍率下进行的。高倍率充放电造成的结果就是电池的SEI变厚、镀锂加剧，活性锂加剧减少，电解液加速分解或者和正负极发生副反应，甚至电池会产气使得一些部件接触不良，导致电池内部的离子阻抗、电子阻抗、欧姆电阻、电解液粘度增大，活性物质和集流体的接触愈加不好。总体而言，就是会使电池的整体内阻增大。根据欧姆定律，Q=I2Rt, 电流和时间在F6也能有所体现，所以该特征重点关注的就是R，而R也能通过F9体现出来。综合F6和F9，就能从理论上算出Q。而实际的放热量Q还需要考虑恒温箱的冷却情况。至于恒温箱的温度情况，这个每组实验都是在30摄氏度温度下进行的，每次都是测40多个电池，电池的型号都是一样的，所以放热量在training set, primary test set, 和secondary test set中应当具有一致性。但此测量参数与其他测量参数的关系相对上面几个（F1,F4,F8）参数并不直接，故F7放在最后去除。