开场：我们工作的主要内容是对模型的进一步优化，以15号风机做训练集，21号风机做测试集，根据测试效果来改进特征和规则。

回顾：

以 15 号风机各变量方差的百分之二十五分位数作为阈值进行过滤，剔除掉方差小于阈值的变量，保留方差大于阈值的变量（21 个），

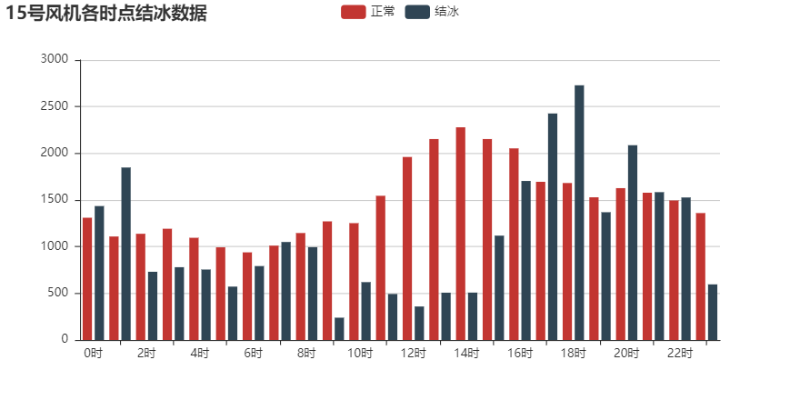
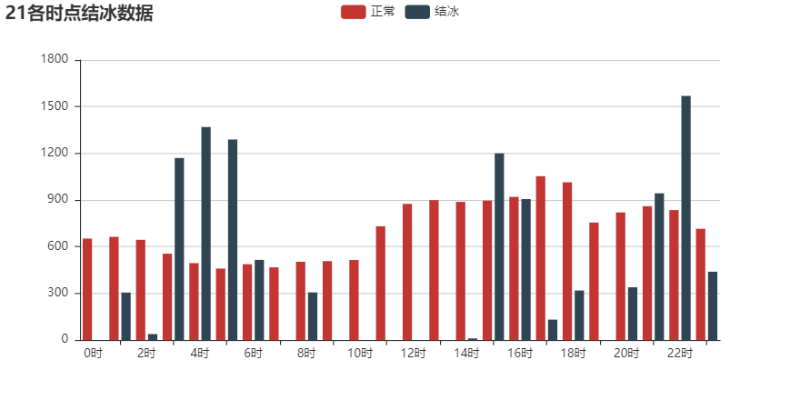
在模型优化的过程中，我们发现如果保留的原始属性太多，再加上一些新构造的物理意义向量（扭矩、cp系数等），很容易导致在21号风机上过拟合。

卡方过滤

对方差过滤后的 15 号风机数据进行归一化，以 0.05 为显著性水平，认为大于显著性水平的变量为相关性低的变量，对其进行剔除，从而将变量个数降低到 16 个。

最后再加上'temp\_diff', 'torque', 'cp', 'ct', 'pitch\_angle\_std', 'spd\_rate', 'pitch\_speed\_ave'物理意义构造的特征变量，共23个。

强规则过滤：对于哪部分的数据需要强制删除



模型效果：

强规则过滤后RF+LR的效果最好

RF效果没有变化

RF+LR的效果比过滤前提升8个点

GBDT和GBDT+LR的效果不如过滤前

数据切割后period=0的部分除GBDT外其余结果都变差

Period=1的部分所有模型效果都变差

不切割而是作为特征学习效果最好