一种二类类别不平衡时序数据分析方法

一. 成果完成人及所在单位：徐华卿等 电气工程与自动化学院 指导教师：彭彦卿

二. 成果简介：

实际生活和生产应用中，已有的机械或电力电子设备的故障率极低，在历史数据中设备正常状态和异常状态的数据分布极端不平衡，这种广泛存在的类别不平衡问题使得很多经典算法有很大的误差。非线性状态估计技术（NSET）可以很好的解决类别不平衡的二分类问题，但是此技术对时序数据没有很好的针对性，本成果结合非线性状态估计技术和常用的时序数据分析方法应用于设备历史数据的分析，精确得出设备异常状态出现的时刻点。实际应用数据为 UPS 设备历史数据。

三. 主要研究内容及创新点：

通过对 UPS 设备的历史数据的分析，找到一种方法能够定位到具体的故障时刻，并且这种方法能够一般化地应用到别的设备历史数据。主要内容：由于实际工程应用中的设备一般不容易发生故障或异常事件，也就是说在历史数据库中正常类别的样本远远多于异常类别的样本，如果把所有异常类样本均视为故障类，那么则是一个二类类别不平衡问题；通过快速傅里叶分析得出适合的样本分组周期为 T，于是找出所有在一个周期内有连续采样间隔的样本，并根据数据库内的日志信息给这些样本贴上类别标签，通常此时的特征空间含有很多无用特征，于是通过顺序后退法（SBS）做特征选择，选择出一个最佳特征后，再用常用的时域特征提取方法在一个周期内从这一个最佳特征中再提取出一些时域特征（如均值，方差，峰度，偏度，峭度等等），这样提取之后的样本间隔为一个周期 T，这样就形成了一个多维时域特征空间，并且时域特征之间相互关联，这样就可以应用非线性状态估计技术了。与通常的做法不同，这里不对选进 NSET 的记忆矩阵样本做复杂处理，直接将类别全部为正常状态的时序连续样本选取进记忆矩阵，进一步地为了避免认知矩阵不可逆，直接引入正则项求其广义逆阵。最后，根据 NSET 的估计值和测试值的相似性度量可以很好地找到包含异常类的周期，进一步地，再在异常周期内如法炮制应用 NSET 就可以找到异常类时刻点。

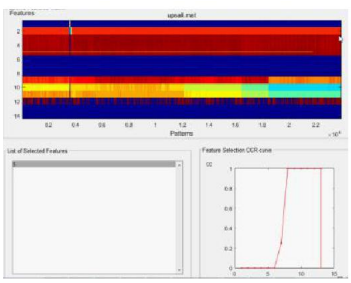


图 1 SBS 特征选择

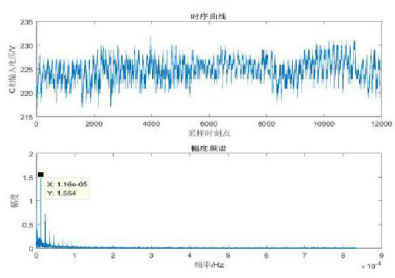


图 2 快速傅里叶分析

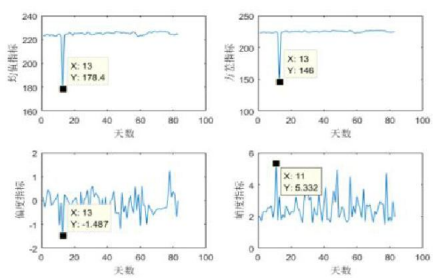


图 3 四个时域特征

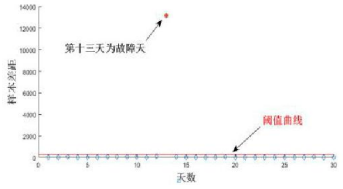


图 4 NSET 找到异类样本周期

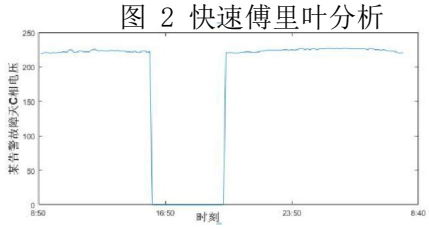


图 5 异类样本周期原始特征

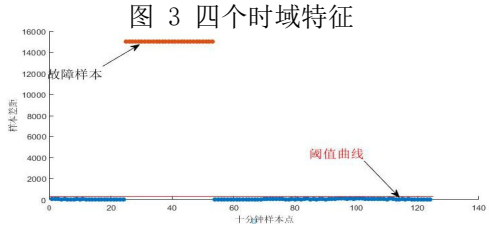


图 6 NSET 找到原始特征异类样本

四. 主要研究成果及应用情况

四项发明专利申请中。

联系人：徐华卿， 电 话： 18959297584 Email:huaqing.xu@qq.com