窗体顶端

赛题题目：空调制造过程中制冷剂灌注量的预测与优化建模

一、背景介绍

空调系统在现代生活中已成为必不可少的设备，广泛应用于家庭、办公楼、工业设施等场所。空调系统的核心在于其制冷循环过程，而冷媒（制冷剂）是实现这一循环的关键物质。冷媒在空调系统中不断循环，通过蒸发和冷凝两个过程吸收和释放热量，从而调节环境温度。冷媒量的准确控制直接影响空调系统的运行效率和可靠性。

传统上，冷媒的灌注通常依赖技师的经验和一些基本的测量工具，如压力表和温度计。这种方法存在显著的局限性，因为经验和简单工具无法充分考虑到环境变化和系统复杂性的影响。冷媒灌注过多会导致系统压力过高，增加压缩机负担，甚至引发设备故障；冷媒不足则会导致制冷效果不佳，系统长时间高负荷运行，增加能耗并缩短设备寿命。

在现代空调系统中，实际冷媒灌注量的测量通常通过质量流量计实现。质量流量计是测量流体质量流量的一种仪表，它通过测量流体在流量计内部的压降和温度等参数，计算出流体的质量流量。质量流量计的输出信号一般为脉冲信号，需要通过转换计算才能得到对应的流量。具体而言，质量流量计的脉冲信号与流体的质量流量成正比，因此可以通过统计单位时间内的脉冲数来间接地计算出流体的质量流量。这种方法不仅提高了测量精度，还简化了操作流程，使冷媒灌注更为精准。

随着科技的进步，数据驱动的智能预测方法开始应用于空调系统中，以提高冷媒灌注的准确性和可靠性。通过在空调系统中安装传感器，实时采集系统运行数据，如温度、压力、运行时间、电流和电压等，利用这些数据构建机器学习或统计模型，可以精确地预测冷媒的最佳灌注量。这种方法不仅能显著提高空调系统的运行效率，减少能源消耗，还能有效延长设备寿命，降低维护成本。此外，冷媒的准确控制还有助于环境保护。许多冷媒对臭氧层有破坏作用，或者是强效的温室气体。通过精确灌注冷媒，减少其泄漏和不必要的排放，可以有效降低对环境的负面影响。这对于推动绿色技术的发展，减少碳足迹具有重要意义。

综上所述，基于数据的冷媒灌注量预测不仅是提升空调系统运行效率和可靠性的重要手段，更是节能减排、保护环境的关键措施。研究和应用这一技术，具有显著的经济效益和社会效益，对行业发展和环境保护都具有深远的影响。

某空调厂家在对空调外机灌注冷媒过程中，首先需要使用真空泵对冷媒罐进行预抽真空，抽空压力<=80pa则进行保压三秒，再次检测压力若<=200pa即可进行灌枪灌注，否则继续进行抽真空。在灌注过程中需要控制冷媒管道的最小压力，同时通过数据回溯可得到样机生产时的电压电流，室温湿度等生产参数。

目前，在冷媒灌注量预测中，为了节约时间和成本，通常采用建立灌注量预测模型来实现灌注量的精准控制：针对灌注机器的运行原理，收集一系列生产过程中与灌注量相关的可测量数据，然后以这些生产数据作为自变量，质量流量计测得的脉冲数为因变量，建立空调冷媒灌注量的定量关系模型，然后使用该模型预测不同生产条件下空调的冷媒实际灌注量，或者指导已有生产参数的结构优化。

二、数据集介绍

本案例针对流量计的脉冲数提供了7000个样机同一灌枪的生产过程数据，其中包含了每个空调样机在生产过程中的12个相关参数数据（自变量），其中f1:为保压三秒之后的检测压力/pa；f2:为真空泵预抽真空状态后的抽空压力/pa；f3:开始灌注时系统设置压力/Mpa（系统设置）；f4：由冷媒管道返回的灌注过程最小压力/Mpa；f5:电压/V（灌注机电压）；f6：电流/A（灌注机电流）；f7：生产线检测时长平均制冷功率/W；（空调制冷平均功率）f8：生产线检测时长平均制热功率/W（空调制热平均功率）；f9：订单批次；f10：厂房环境温度/℃；f11: 室内风速/m/s；f12：室内湿度/%RH；target:质量流量计所接收脉冲量。

三、问题解决

1.根据提供的数据，针对7000个样机的12个生产参数进行变量相关性分析和变量选择，根据变量对脉冲量的影响程度进行特征重要性排序，并给出前三个对脉冲量影响最显著的三个自变量。同时请详细说明变量筛选过程及其合理性。

2.请结合问题1，选择不超过4个生产参数数据构建定量预测模型，请详细叙述建模过程并以指标比较说明模型预测优势。

3.寻找并阐述因变量的哪些生产参数，这些生产参数在什么取值或取值范围时能够使灌注量控制在标准灌注量的附近（此时脉冲量取值范围为：(11570,11600)）。

四、评分标准

1.关键质量特性识别（35%）

（1）数据分析（10%）：剔除无关变量，确保数据的准确性和合理性。

（2）数据预处理（10%）：是否进行异常值和缺失值检查步骤，数据做归一化（标准化）处理。

（3）特征工程（15%）：是否通过合理的方法（如相关性分析、主成分分析等）识别和提取了对冷媒灌注量预测有重要影响的特征。

2.预测模型建立（40%）

（1）模型选择（15%）：是否合理选择了适合的机器学习或统计模型（如线性回归、决策树、随机森林，XGBoost等）并进行原理介绍和描述。

（2）模型训练（15%）：是否对模型进行了有效的训练和参数调整以提高预测精度。

（3）模型评估（10%）：是否使用了科学的评估方法（如交叉验证）对模型进行性能验证，计算并报告了模型的误差指标（如MAE、RMSE、R²等），并达到了较好的效果。

3. 确定自变量最佳取值范围（15%）

（1）方法选择（5%）：是否选择了合适的方法（如优化算法、敏感性分析等）来确定自变量的最佳取值范围。

（2）分析结果（10%）：是否合理解释了分析结果，明确指出了各自变量的最佳取值范围。

4. 处理问题过程的完整性（10%）

（1）过程描述（5%）：是否清晰、完整地描述了整个问题解决过程，从数据处理、特征识别、模型建立到自变量取值范围的确定。

（2）成果展示（5%）：是否形成了一份详细的报告，包含数据处理过程、模型选择理由、预测结果、自变量最佳取值范围的确定过程和改进建议。

窗体底端