

综合实验题目二

管线钢性能预测与因果推断

1 问题背景

管线钢（如图1所示）是广泛应用于石油和天然气长输管道的核心材料，其力学性能要求极为严格，需要在保证高强度的同时兼具良好的韧性和延展性。随着钢材强度等级的不断提高，传统基于试错法的工艺调试手段逐渐显现出效率低、成本高、性能波动大等问题。



Figure 1: X80级管线钢

当前工业生产过程中，管线钢性能的不确定性主要体现在以下方面：

- 批次之间的性能波动（不同炉次、不同轧制条件）；
- 组织特征与工艺条件之间存在复杂的非线性关系，难以精准建模。

为了解决上述问题，钢厂逐步引入大数据与机器学习手段，期望通过机理模型与数据驱动算法融合方式，开发兼具预测能力与因果分析功能的统一建模框架，从而实现：

- 针对制造过程数据（如化学成分和工艺条件），预测管线钢的多项性能指标；

- 对影响管线钢质量的关键因素进行识别与分析。

本题目即以某钢厂提供的管线钢性能真实数据（“x80钢种数据.csv”）为基础，结合相关领域知识（“数据说明.xlsx”），开展管线钢性能预测与因果推断相关研究任务。

2 数据说明

在管线钢的生产过程中，通常经历“加热—轧制—冷却—取样检测”等关键工序（如图2所示）。加热阶段涉及升温曲线、保温时间和出炉温度等参数，会影响奥氏体（Austenite，是钢铁的一种层片状的显微组织）晶粒的大小，进而影响钢材性能；轧制阶段包含轧制温度、变形量、轧制速度等条件，决定了动态或静态再结晶过程及晶粒演变；冷却阶段则通过控制冷却速率和终冷温度，调节铁素体、珠光体、贝氏体等相的转变，最终影响管线钢的显微组织。经过上述工艺后，会对材料进行取样和性能测试，得到屈服强度、抗拉强度、延伸率、冲击功、DWTT 剪切面积等力学性能指标。

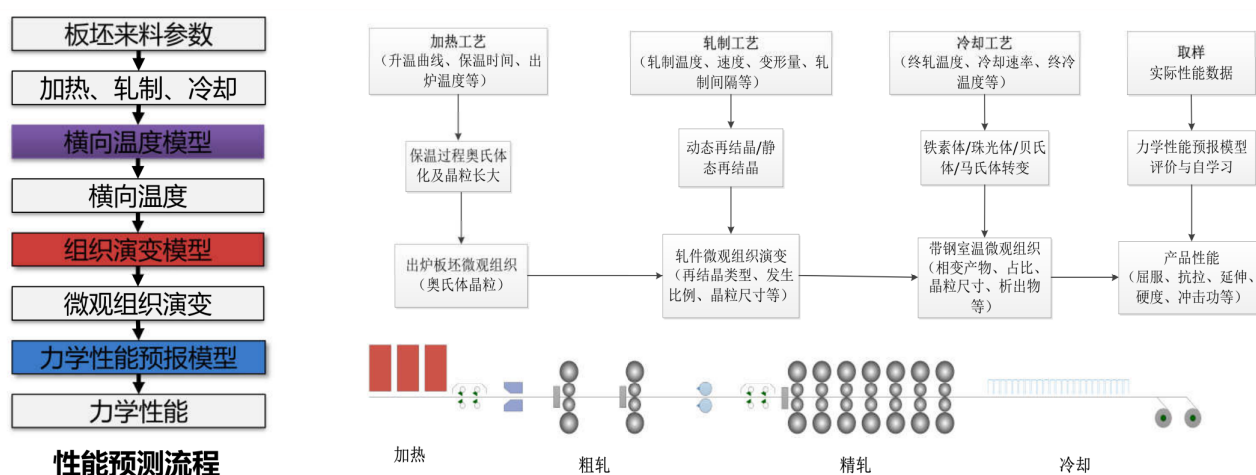


Figure 2: 管线钢生产与性能预测流程图

在现代化的数字化生产环境下，这些工序参数及对应的组织、性能数据均能被完整记录，构成了一个覆盖从原料、工艺到性能的多维数据库。本题所提供的管线钢性能数据库正是基于上述流程中采集的真实生产数据，涵盖了化学成分、加热制度、轧制参数、冷却条件以及最终的性能测试结果。

本题聚焦于力学性能预测模型（温度模型、演变模型与本题无关），以“x80钢种数据.csv”为工业生产原始数据，进行数据模型建立与分析。“x80钢种数据.csv”以试样板坯号（列名：SLAB_NO）为索引，记录了近20000次测试中管线钢生产流程各项参数以及质量性能响应，对应列名解释以及详细相关领域知识可参考“数据说明.xlsx”。

2.1 目标变量

本题中需要分别建模并分析以下3种力学性能指标:

指标中文名	列名	单位	含义说明
抗拉强度	RM	MPa	材料最大承载应力
冲击功	IMPACT_RST_AVE	J	常温或低温冲击试验能量平均值
落锤剪切面积	DWTT_AVE	mm ²	DWTT试验中剪切面积的平均值

Table 1: 3种主要钢铁性能指标

2.2 特征变量说明（部分）

生产中采集的特征变量，主要包括钢材基础信息、化学成分、“加热—轧制—冷却”生产流程参数以及其他信息。部分特征变量如下表所示:

	列名	中文解释
基础信息	STEELGRADE	轧制钢种
	THK	订单钢厚度
成分	C_CHEM_VALUE	碳含量
	MN_CHEM_VALUE	锰含量
	SI_CHEM_VALUE	硅含量
加热工序	CHARGE_TEMP	装炉温度
	EXT_TEMP	出炉温度
	RHF_DUR_TIME	炉内驻留时间
轧制工序	SLAB_THK	铸坯厚度
	CR1_REDUCT_TEMP	精轧开轧温度
	MILL_END_AVE_TEMP	终轧平均温度
冷却工序	TEMP_WATER	水温
	PRESSURE_B	水压
	SPEED	速度

Table 2: 部分特征变量对照表

完整变量定义、含义以及特征机理，请参考“管线钢性能预测信息.xlsx”文件。

3 实验任务

● 任务一：特征工程与变量筛选

- 1.1 对原始变量进行可视化与统计描述，识别异常值，并做数据预处理；
- 1.2 分析变量间的相关性，结合共线性指标、信息增益等方法进行变量筛选；

- 任务二：建模与性能预测

- 2.1 对3项性能指标分别构建机器/深度学习模型，训练集和预测集比例为8 : 2;
- 2.2 建议比较多种建模算法（例如线性回归、随机森林、XGBoost、LGBM、MLP等）；
- 2.3 输出模型的评价指标（如MAE、RMSE、 R^2 ）。

- 任务三：因果关系挖掘

- 3.1 利用SHAP（[GitHub链接](#)）对模型特征重要性进行解释性分析；
- 3.2 尝试采用因果发现算法（如DoWhy[arxiv链接](#), PC算法[arxiv链接](#)，或其他可行方法[arxiv链接](#)）构建候选因果网络；
- 3.3 分析主要工艺变量对抗拉强度、冲击功以及DWTT的影响路径，提出“合理假设解释”。