



清华大学

Tsinghua University

工业大数据分析指南解读

王晨

联盟工业大数据特设组

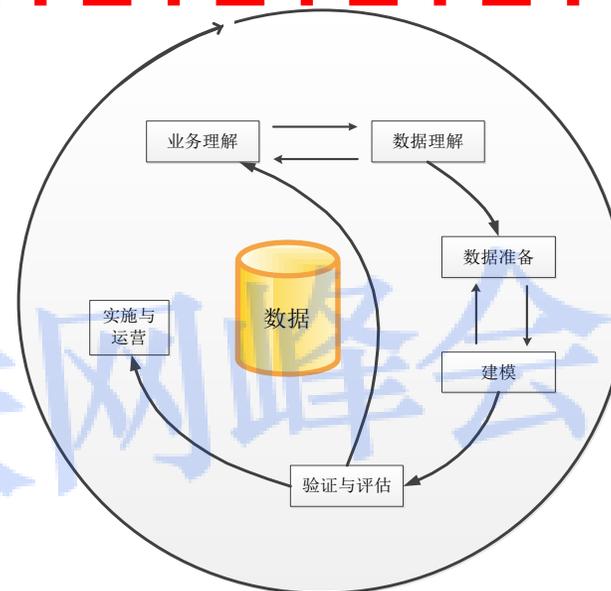
清华大学 大数据系统软件国家工程实验室

在工业互联网产业联盟的指导下，工业大数据特设组主持编写了《工业大数据分析指南》

指南全文架构

一	工业大数据分析概论
二	工业大数据分析框架
三	业务理解
四	数据理解
五	数据准备
六	数据建模
七	模型的验证与评估
八	模型的部署
九	展望未来

六个基本步骤



- **CRISP-DM模型**

工业大数据分析概论

工业大数据分析概述

- 工业大数据分析的概念
- 工业大数据分析的相关技术
- 工业大数据分析的基本过程
- 工业大数据分析的类型
- 工业大数据分析的价值
- 工业大数据分析支撑业务创新

工业大数据分析的特殊性

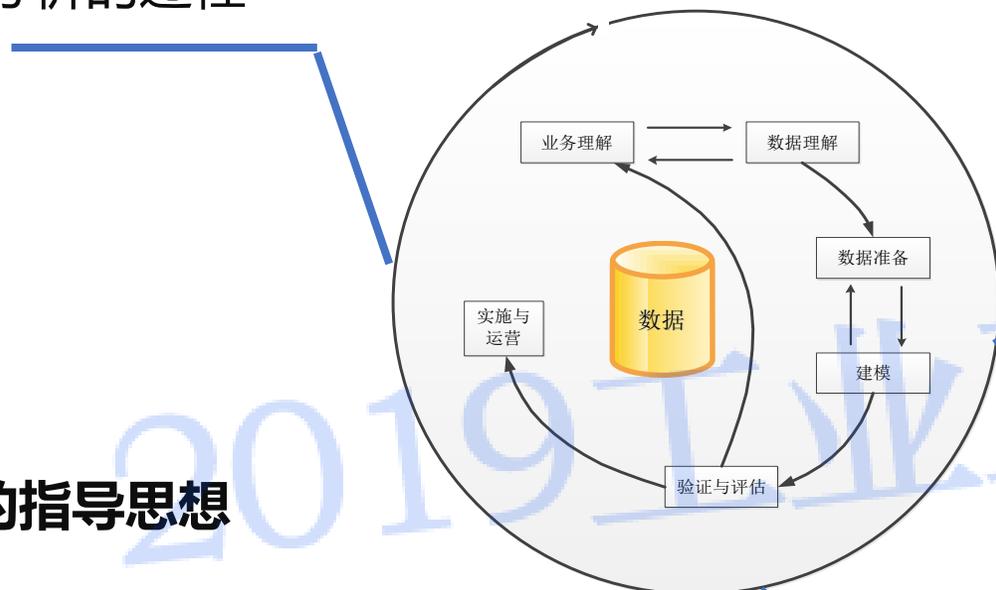
- 从工业数据分析到工业大数据分析
- 工业大数据与商务大数据分析
- 工业大数据建模的难点

工业大数据分析的常见问题

- 业务和数据理解不当导致的失误
- 建模和验证过程的失误
- 避免失误的方法

CRISP-DM模型（跨行业数据挖掘标准流程）

以数据为中心，将相关工作分成业务理解、数据理解、数据准备、建模、验证与评估、实施与运行等六个基本的步骤，用以指导工业大数据分析的过程



CRISP-DM模型落地难点

工业数据关联关系复杂、工业数据质量差、工业场景的分析要求高

工业大数据分析的指导思想

用好CRISP-DM，减少不必要的反复，提高数据分析的效率

认识工业对象

用系统的观点认识工业对象，理解系统功能并进行功能描述，深入认识系统的功能原理、注重系统功能与业务场景的关联

理解数据分析的需求

工业过程中的数据分析需求：DMAIC模型，界定D、测量M、分析A、改进I、控制C等五个步骤
数据分析的价值需求：确认两方面问题：价值是真的存在？价值是否足够大？
具体业务场景的数据分析需求：不同场景下问题之间的关系会发生改变，分析和应用结合流程
数据分析需求的梳理方法：用5W1H方法深入理解需求、将业务目标相关的因素进行分类

工业数据分析目标的评估

理解工业知识、分析工业知识的合用性、专业领域知识和数据模型融合

制造的全生命周期

跨生命周期数据的管理和分析

四

数据理解

数据来源

业务与数据的关系
离散行业的数据源
流程行业的数据源

数据的分类及相互关系

工业数据的分类
数据间的关联关系

数据质量

数据质量的定义
数据质量的组成要素
数据质量的影响因素

五

数据准备

不同业务系统之间的数据整合和共享，打破“信息孤岛”，拆除“数据烟囱”，实现多源基础数据的按需互通和共享。

针对要解决的问题开展数据治理，实现数据资源的互通和共享



基于统一的标识解析体系实现数据的互联、互通、共享和溯源

- 数据预处理
- 数据异常处理
- 数据缺失处理
- 数据规约处理

1

模型的形式化描述

基本描述

模型的深入表述

对建模思想的影响

2

工业建模的基本过程

建模的基本思路

模型融合的方法

模型的优化过程

3

工业建模的特征工程

数据初步筛选

特征变换

特征组合

特征筛选

特征的迭代

4

工业数据分析算法介绍

传统统计分析类

通用机器学习类

旋转设备振动分析类

时序数据时间序列类

非结构化数据文本挖掘类

统计质量控制类

排程优化类

七

模型的验证与评估

知识的质量

- 知识的确定性与准确性
- 知识的使用范围
- 知识的质量与可靠性

传统数据分析方法及问题

- 基于精度的验证方法
- 精度验证方法的局限性
- 解决验证问题的传统方法

基于领域知识的模型 验证与评估

- 对适用范围的评估
- 对精度的评估
- 场景的综合评估
- 模型的迭代评估

八

模型的部署

模型的部署

模型部署前 应考虑的问题

模型部署对工作方式的改变
模型部署的标准化与流程化
模型部署的自动化与智能化

实施和运行中的 问题

数据质量问题

运行环境问题

精度劣化问题

范围变化问题

问题的解决方法

- 改善数据收集、限制应用范围
- 实时性（分布式消息队列和流处理）、计算效率（近似算法、并行算法和流式算法）、存储量的可扩展性（分布式系统）、计算稳定性（集群计算框架）
- 采用本质性的关联；修正模型
- 模型的应用限制在特殊范围内

部署后的 持续优化

- 长时间的优化改进
- 提高维护和管理水平
- 扩大模型适用范围
- 适应新知识

九

展望未来



A

工业大数据逐渐成为传统制造业与新一代信息技术深度融合的落脚点



B

政府的政策制定、驱动导向、发展引导、配套支持等，为工业大数据产业发展创造了优良环境。



C

工业大数据技术创新正逐渐从技术驱动转向应用驱动，广阔的市场空间和大量的应用需求为工业大数据发展提供了强大的驱动力。



D

工业大数据将成为推动制造业创新发展的重要基础，为中国的工业升级和转型注入强大动力。



E

工业大数据必将迎来高速发展的历史阶段。



谢谢聆听！

序：谢少锋

编写组成员：王建民、郭朝晖、王晨（清华大学）