

# Causality , Correlation and Messy to the Industrial Big Data

**Zhahiu Guo**  
**BaoSteel LTD.**

# 可靠度要求及因果性

## Reliability Requirements and Causation



高价值和高可靠性要求往往是硬币的两面

如果正确判断能带来巨大效益，  
错误判断也可能带来巨大损失。

工业界对分析结果的可靠性要求很高。



$$y = bx, y = b(x + \eta) + \xi$$

$$E(\hat{b}) = b \frac{Dx}{Dx + D\eta}$$

“有偏估计”典型值可达50%~70%，后果是无法外推。

这可能使得各种基于“误差最小”的算法失效。



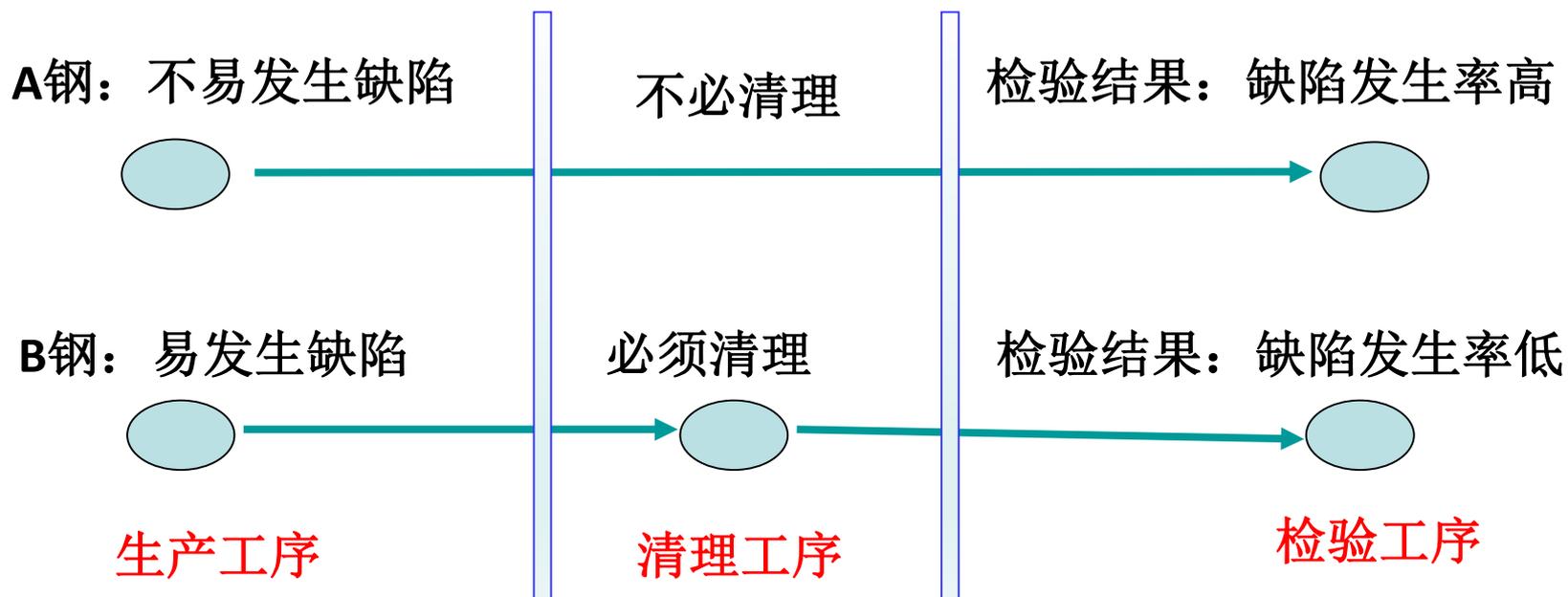
练太极拳的，身体差

如果止步于相关分析  
多数结论是错的和无聊的

系统分析一定要重因果  
否则很难达到可靠的要求



杀了公鸡，太阳照样升起



系统性、多维度和领域知识  
Systematicness 、 MultiDimensional  
and the Domain Knowledge

隐性的、非随机干扰多

分析缺陷与工艺参数的关系

每年发生率不同。  
与钢种有关  
    固定钢种：  
与精整有关  
    再固定精整：  
与宽度有关  
    再固定宽度：  
与检查人员相关。  
    再固定检查人员：  
与正反面相关。  
.....

难以达到“频度稳定”的基本要求

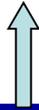
大量系统性干扰：数据需求量上升

对于可靠性的取得，  
数据量总是不够的。

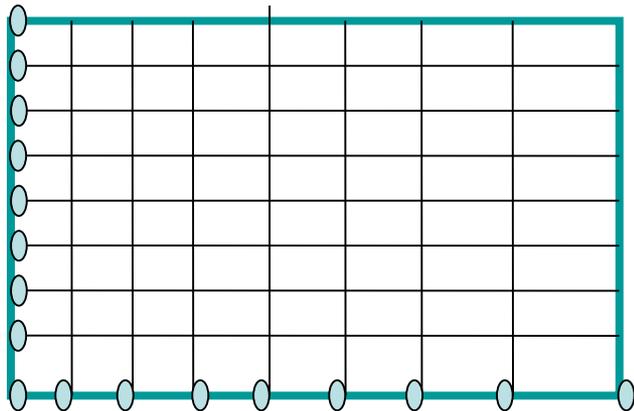
To Obtain Reliable Results,  
The Amount of Data is Always Insufficient.

## 基于领域知识的知识发现

$$f(x,y) = g(x) + h(y)$$



数据需求量由 $n^2$ 变成 $2n$



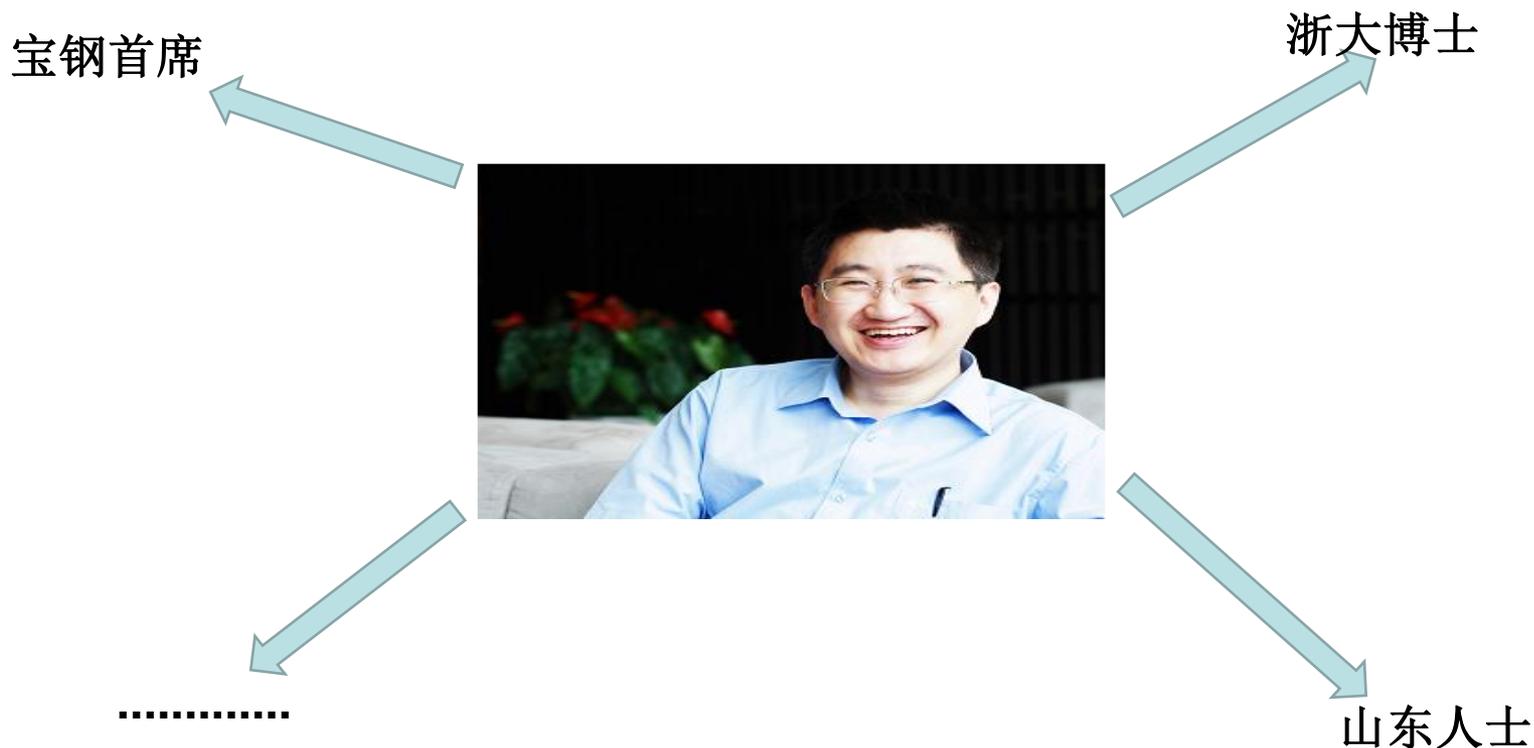
要得到可靠的结果，单靠数据本身永远都不够：如同人无法举起一座山。不能指望数据有遍历各种工况的“全集”。

领域知识让我们减少了对数据的需求量，让知识发现变得可能。

**领域知识：降低对数据量的需求**

混杂性：提高可靠的程度

Messy : Improve the Reliability

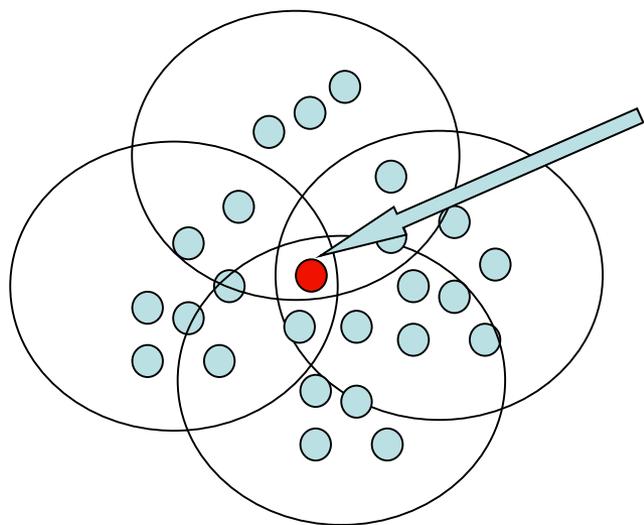


混杂，意味着信息独立；  
精确都的上升（山东、济南、历城），信息编辑增量的递减.....

- 更多的证据。
- 独立性强的证据。
- 更可靠的证据。
- 更严密的证据链。
- 没有明显的反例。
- 理论佐证与相互竞争。

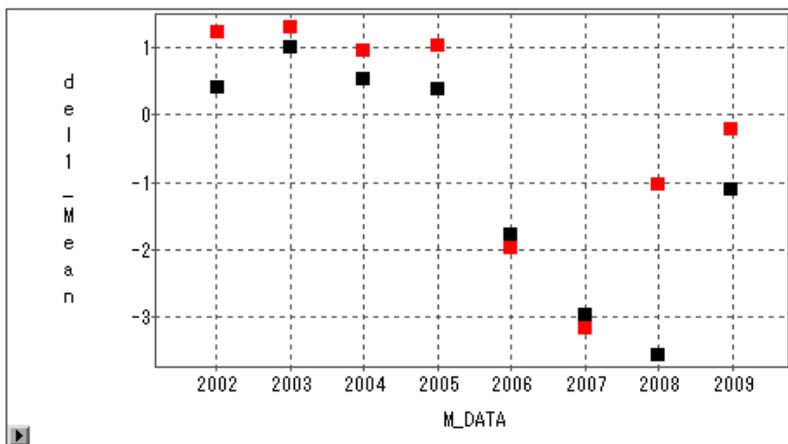


元芳，你怎么看？

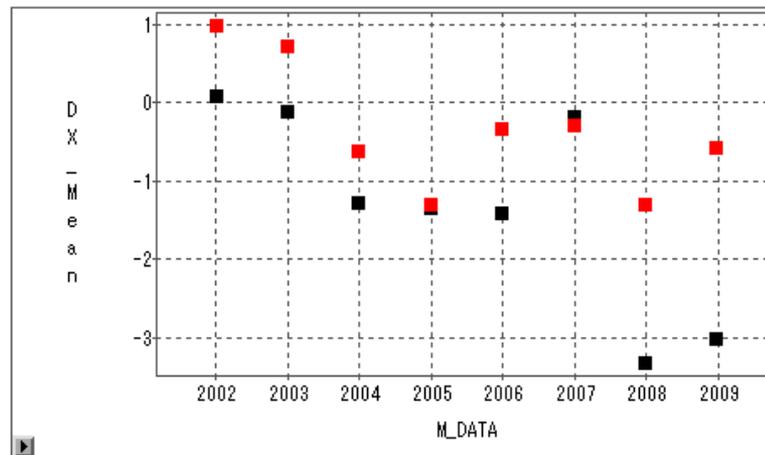


认定的结论

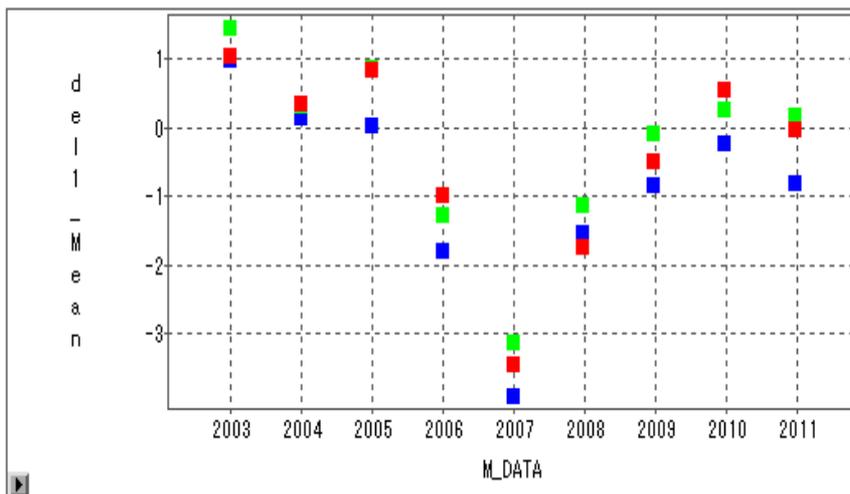
每一个独立的证据，都可能会排斥或者加强对一个结论的认识



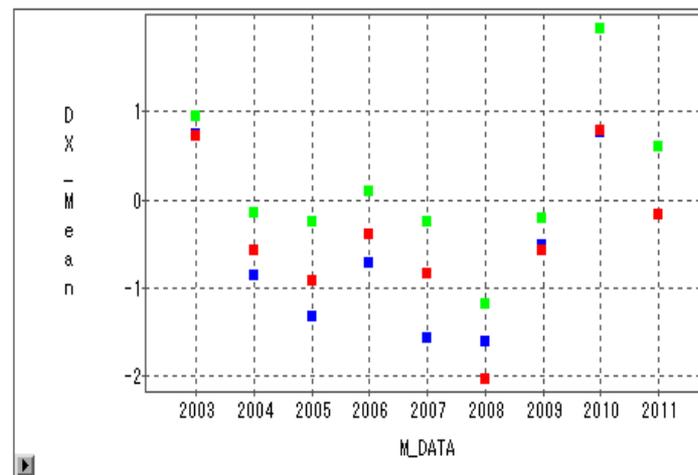
塑性 & 液位波动



韧性 & 液位波动



塑性 & 板坯位置



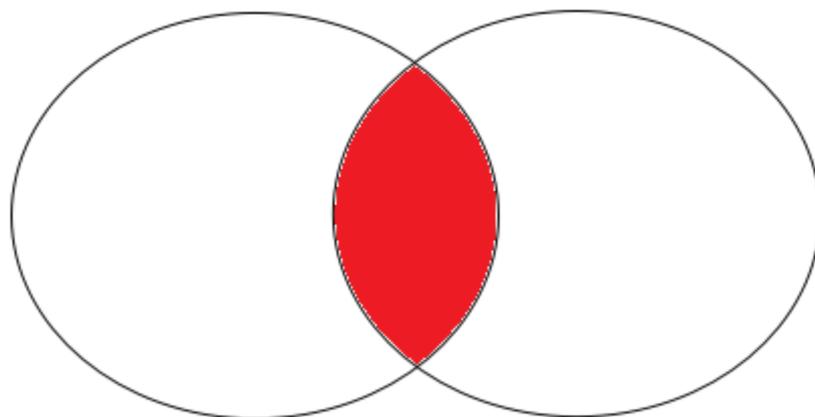
韧性 & 板坯位置

因果的取得：领域知识的融入

Causality Achievement:

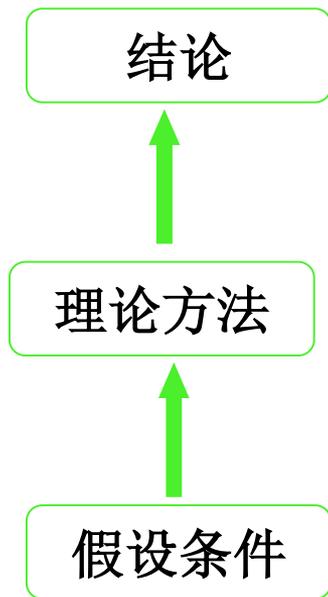
Relying on Domain Knowledge

“可靠”的结论，是人的领域知识和数据分析结果所形成的共识

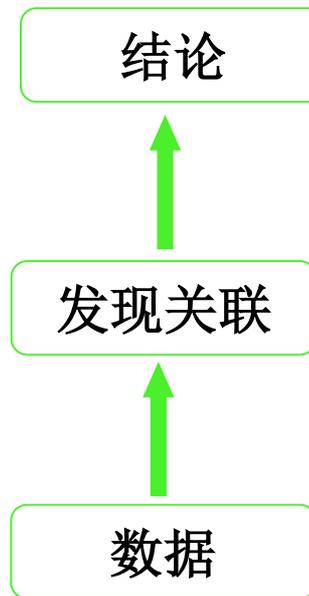


领域可解释是前提：融入因果

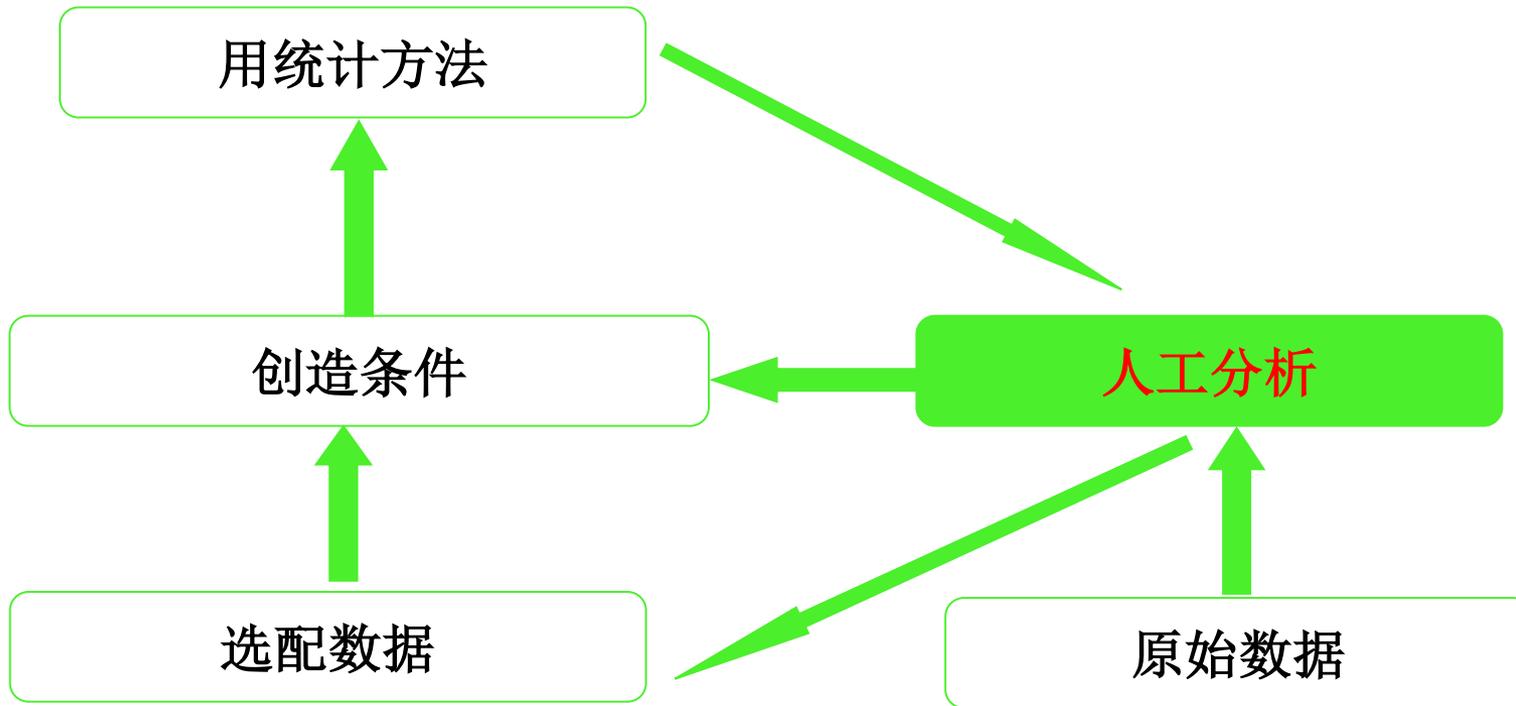
数据分析和领域知识的关系



传统统计学：  
从假设条件出发



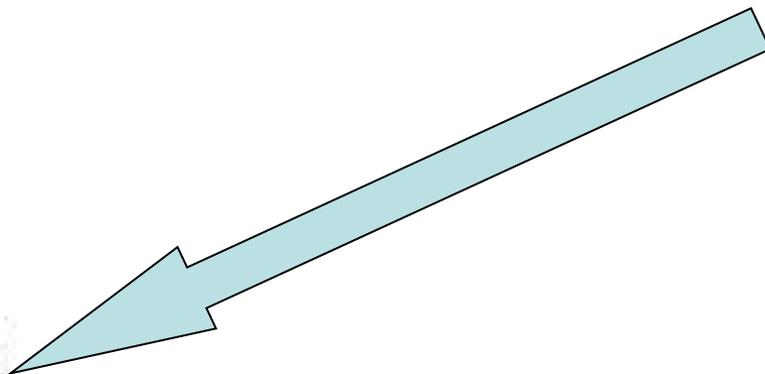
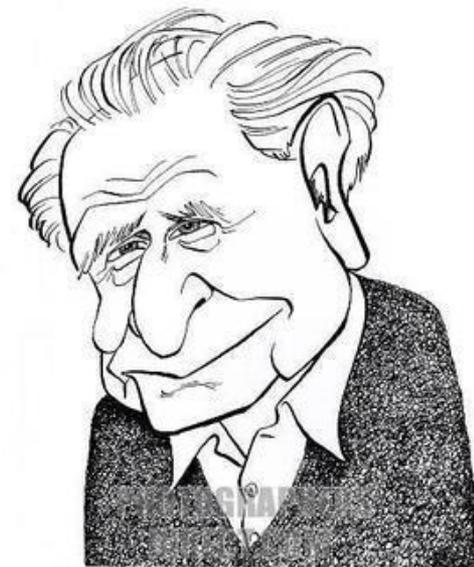
数据挖掘：  
不预设前提条件



统计和机理的结合

对条件的判断常常来自于数据之外的认知

数学规律： 已知的正确，未知的也正确。  
物理规律： 已知的都正确，且竞争成功。  
生物规律： 80%的正确。  
经济规律： 50%的正确  
社会规律:.....



证明数学定理，一个证明就够了，  
证明历史事件，至少要5个证据

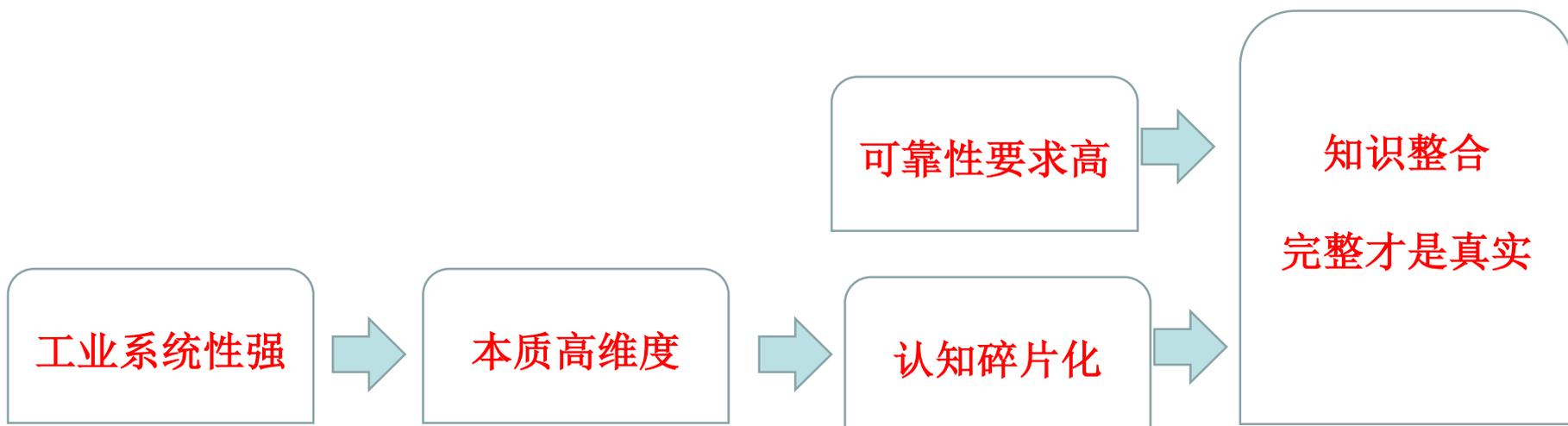
绝对的可靠性：必须放弃的奢望



把数据中的现象、对机理的认识融合起来



从碎片化到完整认识



工业互联网：看到了几层？  
Industrial Internet: How Many Layers Have You Seen?



看到设备诊断，  
看到知识和数据的共享，  
看到制造到服务的转型；  
看到云平台、工业软件，  
看到制造企业向软件企业的转型。

要看到一系列的逻辑驱动。



# Thanks

不必把因果和相关对立起来  
用混杂性将其联系起来。

微信: [guoguo1968](https://www.guoguo1968.com)