# COMP9414：人工智能第6c讲。数据科学与伦理

韦恩-沃布克
电邮：w.wobcke@unsw.edu.au





## 概述

- 问题
  - 过度拟合
  - 偏见和歧视
- **方法**论
  - 特征工程





## 数据科学不是什么（讽刺）。

- 选择一个复杂的概念/统计/指标来衡量
  - 贫困/财富指标，粮食安全地图
- 选择一些大型的数据集
  - 移动电话数据、卫星数据、管理数据、调查数据
- 除了选择一些 "协变量 "之外
  - 夜间灯光，土地使用，等等。
- 将所有数据扔进R/Python的标准方法中，---
  决策树、随机森林、XGBoost、神经网络，---
- 给出了混合的结果（在验证的范围内……）





## 问题：过度拟合

- 地方背景假设
- 汇总和分解数据集
- 验证

过度拟合 =

过于贴合给定的数据，在其他情况下不起

作用

例子。如何不测量财富指数（Blumenstock

等人，2015）。

- 具有5088个特征和856个标记实例的手机数据
- 根据整个数据集（而不是训练集）选择特征
- 不要考虑这个数据的卢旺达特色是什么
- 使用来自另一篇论文的非标准方法
- 忽略合理的（人类产生的）基线
- **5倍交叉**验证产生**5个**

模型，而不是一个索赔（

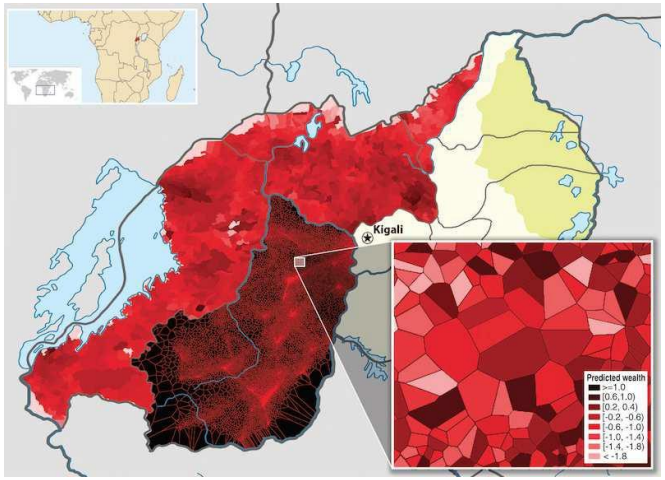
？）许多神经网络/深度学

习模型过度拟合



新南威
尔士大
学
©W.Wobcke等人，2019-
2022年

# 过度拟合








# 问题：偏见和歧视

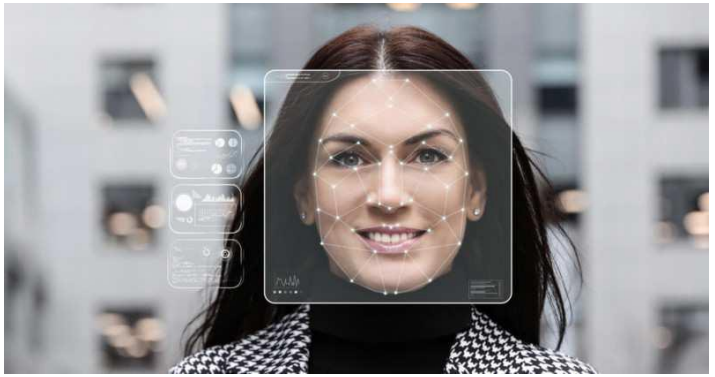
偏见=方法的概括倾向（好的或坏的）。

- **数据集不代表人口**
  - 只有在有电话塔的地区的人才有电话
  - 只有识字的人才能发送短信
  - 只有较穷的人需要 "获得 "电话信贷
- 学习者归纳出 "错误 "的特征
  - 白色背景（只有雪豹的照片是在冬季拍摄的）。
- 学习者 "错过 "了相关特征





# 清华大学







# 面部识别 偏见

- 人口流动的季节性影响（食物短缺）

偏见（在机器学习中）可能导致（不道德的）歧视



新南威
尔士大
学
©W.Wobcke等人，2019-
2022年

# 面部识别偏见

| Gender Classifier | Darker Male | Darker Female | Lighter Male | Lighter Female | Largest Gap |
|---|---|---|---|---|---|
| Microsoft | 94.0% | 79.2% | 100% | 98.3% | 20.8% |
| FACE++ | 99.3% | 65.5% | 99.2% | 94.0% | 33.8% |
| IBM | 88.0% | 65.3% | 99.7% | 92.9% | 34.4% |






# 英国护照歧视





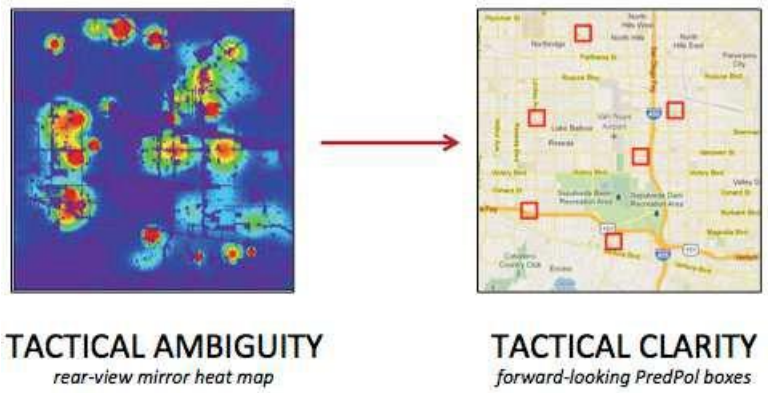


# 错误逮捕的歧视






# 预测性警务歧视



# 累犯评级歧视






# 数据科学方法论

- 方法论。在统计学/机器学习的教科书中
  方法、模型、定理、估算器、技术、工具
- 元方法学。支持这一点的知识和实践
  - 如何决定要测量哪些 "概念"？
  - 如何决定如何定义这些概念？
  - 如何决定如何衡量这些概念（什么数据）？
  - 如何检查结果的稳健性或可靠性？
  - 如何验证结果（内部和外部）？
  - 结果如何影响政策/决策？





# 数据科学的人文要素

当数据的质量、数量有限时，是必不可少的（大部分时间）。

- 人类建议的相关特征
  如果抗议的地点是私人的，那么抗议就不太可能是暴力的。
  - 与冲突进展有关的事件的AfPak本体论
- 人类定义了有用的指标
  - 如果市场在晚上开放，村庄就安全了。
- 人类验证了模型的输出
  - 在15%的随机样本上检查与模型的一致性
  - 验证模型所使用的主要特征
  - 确定比较业绩的基线
  - 与其他数据集交叉检查模型输出





# 特色工程

教科书中缺乏强调，但学习起来非常重要

例子。移动电话数据包括手机信号塔的位置

- 地点是吴哥窟，时间是1天⇒游客？
- 或者，"类似于 "典型的旅游旅行的旅程 ⇒ 旅游
- 地点是购物中心⇒购物（如果不是在家）？
- 最常被叫的人⇒配偶？（如果已婚）。
- 配偶⇒异性(作为支票使用)
- 地点是港口和卡车司机⇒装运
- 卡车的目的地⇒货物的类型？

方法学。强调处理多层次的不确定性



新南威尔士大学
©W.Wobcke等人，2019-2022年

# 当地的背景假设

食品消费得分

- 通过对食物类型进行加权估计，每天2100卡路里
- 权重得到激励，但石油和糖 "需要调整"
- 当地验证（季节性影响，当地变化）。
  北苏丹对南苏丹
  - 喀麦隆的季节性变化
- 与其他措施（行政数据、调查）相关联

理想情况下，衡量能力（？
即使有大量的数据也不可能学会，需要专业知识






# 合并数据集

只使用一种类型的数据对许多目的来说是不够的

- 特别是社交媒体数据（Twitter、Facebook）。
- 特别是在复杂的指标和指数方面
  - 利用医院停车场的图像进行人口保健
  - 使用卫星数据的降雨地点和数量
- 需要三角测量/确证，而不是增加不确定性
  - 需要将独立的数据源 "关联 "起来






# 管道化进程

- 亚洲开发银行的贫困地图（土地使用→回归）。
- 第一阶段的错误很可能是系统性的，而不是随机的。
  - 高斯-马尔科夫假设不成立
  - 需要根据经验估计而不是使用理论
  - 依赖于 "地面实况 "数据集
- 方法与模式
  - 对菲律宾有效（更好），对泰国无效：为什么？
  - 权衡方法的通用性和 "局部验证 "的问题

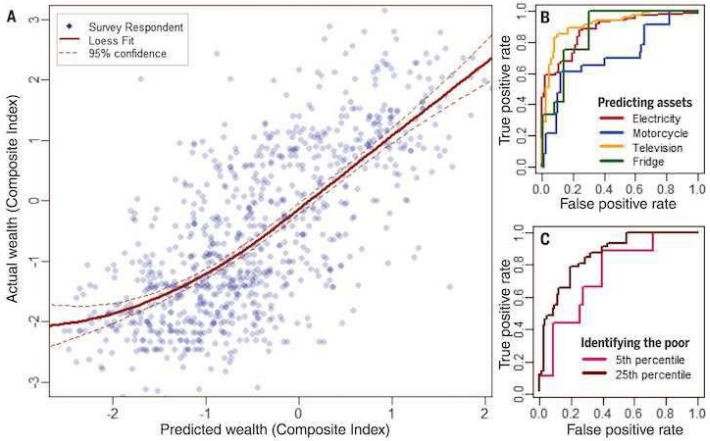





# 切片和切块

- 数据可能只在某些情况下是可靠的
  - 可能能够确定事件的发生，而不是细节
  - 情绪分析出了名的不准确
- 可能想按地区、地位等分析子组。
  - "大数据 "很快会变成 "小数据"
  - 需要统计方法来评估可靠性
  - 将数据的质量映射到所产生的决策的质量上

# 审定









# 总结

数据是否适用于（什么）目的？

- **没有哪个模型是完美的（尤其是学**习型模型）。
- 统计学上的关联性通常非常弱
- **根据当地情况**对模型进行情景化处理
- 与其他数据集交叉检查模型输出
- 表达与结论/决定相关的不确定性
- **"大数据 "方法可以提供 "早期**预警 "信号
- 用不同的时间尺度来补充传统的措施

- 随着假设的变化，不断地验证模型