
ArchData
技术峰会北京站
主办方：
中生代技术
FRESHMAN TECHNOLOGY
快CTO
互联网创业技术服务平台
2017年9月24日北京海淀区丹棱街5号微软亚太研发中心一号楼一层 故宫会议室

TOWARDS AI FOR EVERYONE

第四范式经验与思考分享

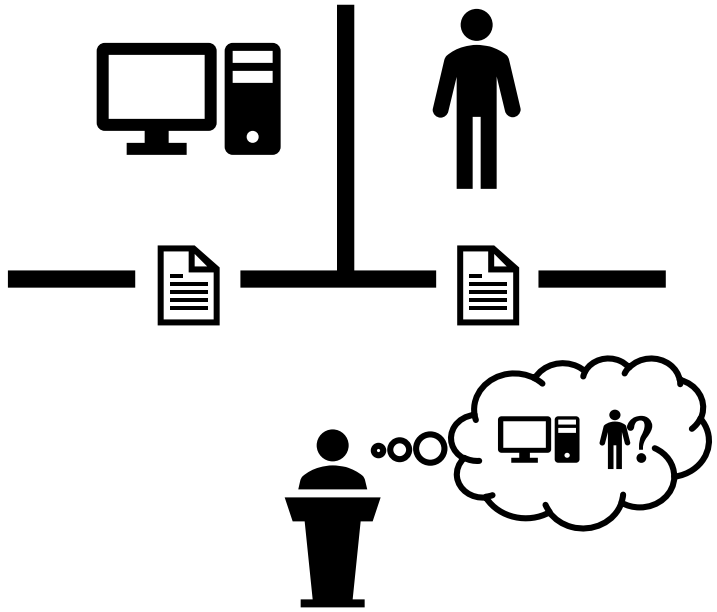
涂威威

目录

- **AI应用背景介绍**
- AI for everyone
  - 效果
  - 成本
- 总结

从图灵测试说起

- 目标：判断机器是否表现出与人等价或无法区分的智能
- 两个基本问题：
  - 充分性：通过图灵测试就是智能?
  - 必要性：通过图灵测试才是智能?
- 两个著名变种：
  - Feigenbaum test
  - Nicholas Negroponte Test



[Alan Turing]

“人工” 智能发展历史

推理期

- 1956-1960s
- 逻辑推理
- 举例：自动定理证明系统

知识期

- 1970s-1980s
- 知识工程
- 举例：专家系统

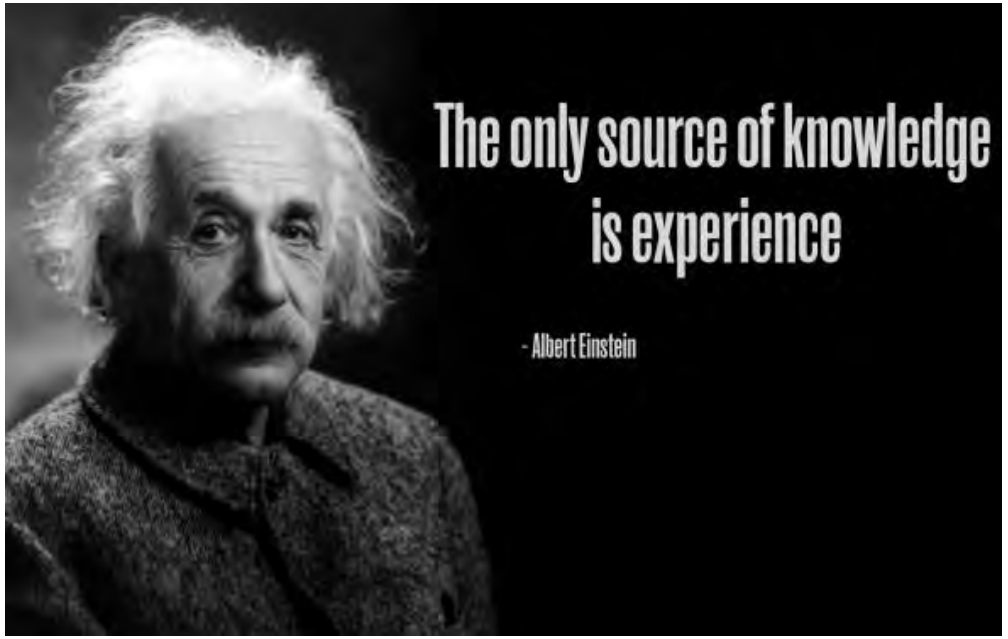
学习期

- 1990s-现在
- 机器学习
- 举例：AlphaGo

[Zhi-Hua Zhou]

机器学习的经典定义

- 利用经验改善系统性能
- 经验 → 数据
- 机器学习被广泛应用
  - 搜索与推荐
  - 生物特征识别
  - 自动驾驶
  - 军事决策助手（DARPA）
  - …



机器学习的成功应用和成本

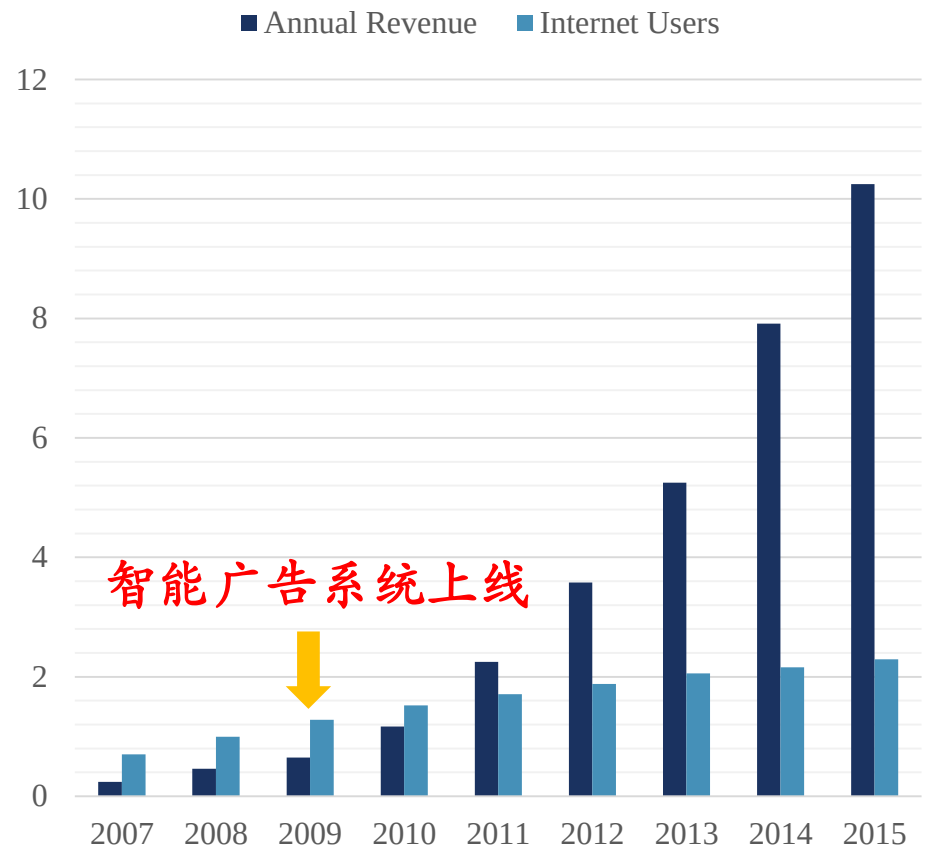
成就

- 战胜人类围棋世界冠军
  - 柯洁、李世石

成本

- DeepMind顶级科学家团队
- 10年以上研究
- ~2000 CPUs + ~300GPUs

机器学习的成功应用和成本



提升

- 虽然互联网用户增长缓慢
- 2009-2015收入提升10倍+

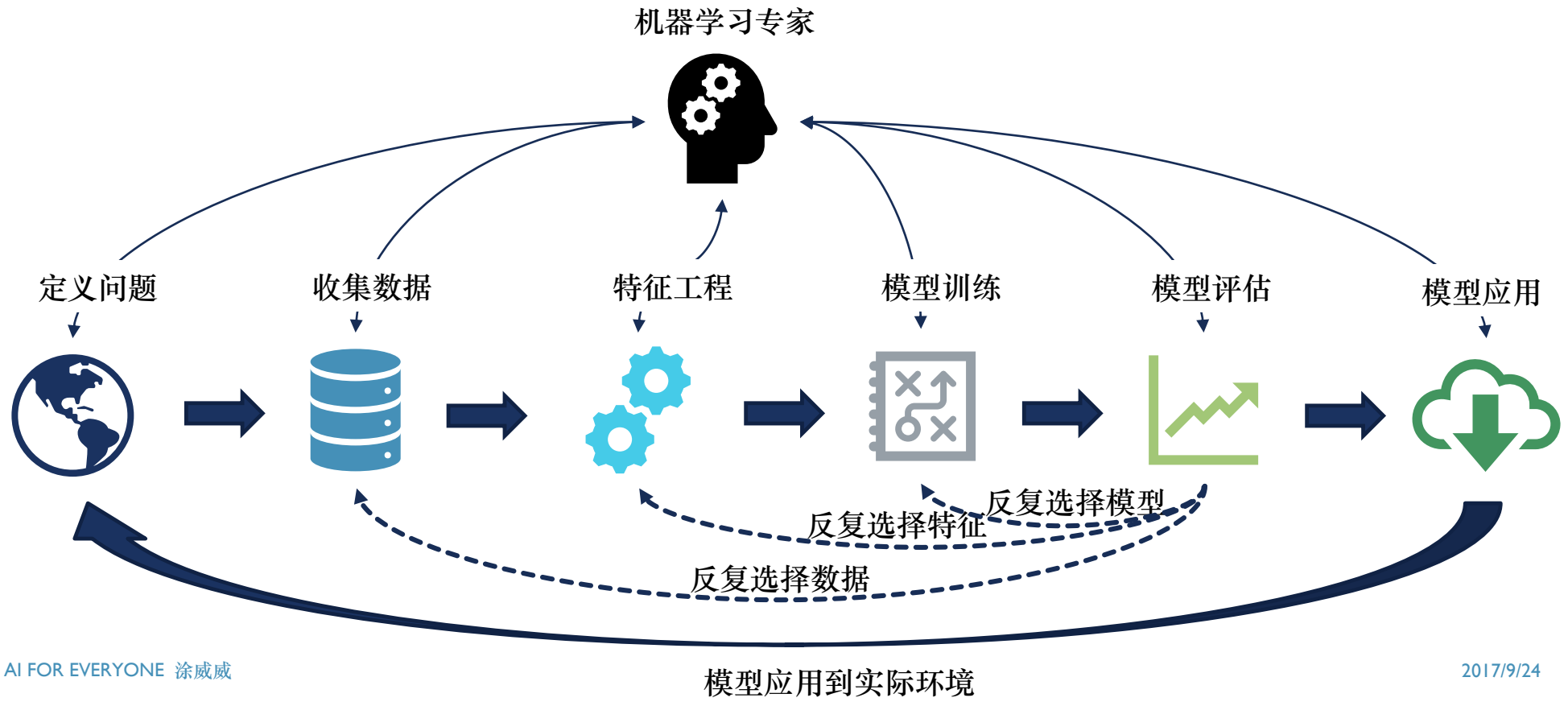
成本

- 线上线下数万台机器
- 60+机器学习专家团队
- 近十年分析调研

目录

- AI应用背景介绍
- **AI for everyone**
  - **效果**
  - 成本
- 总结

典型的机器学习过程

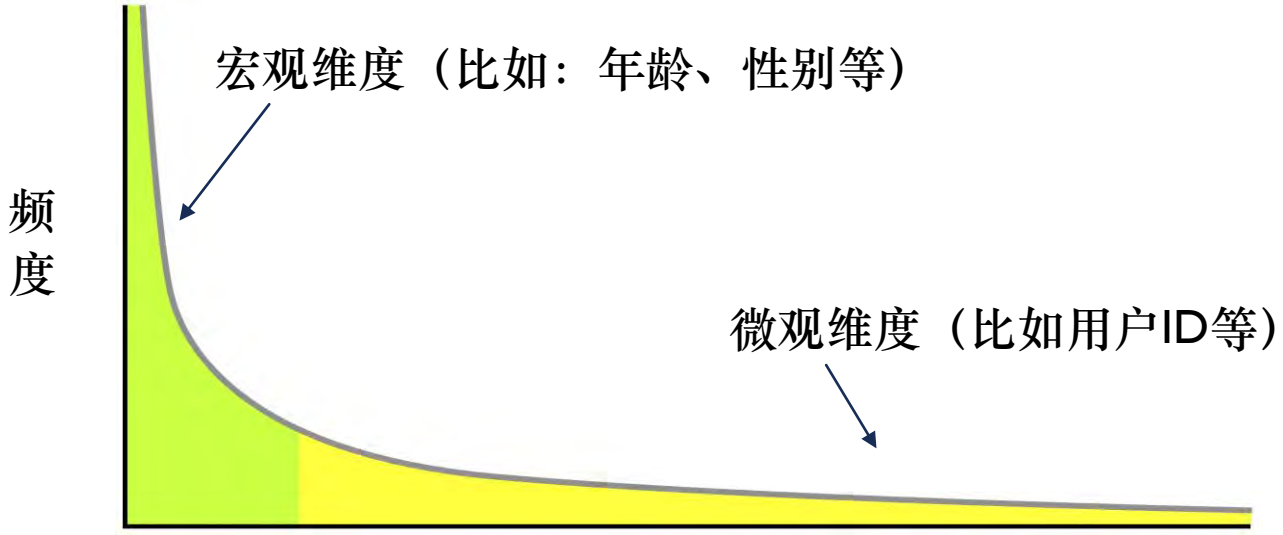

机器学习专家
定义问题
收集数据
特征工程
模型训练
模型评估
模型应用
反复选择模型
反复选择特征
反复选择数据
模型应用到实际环境

机器学习的效果门槛

- 建模门槛
  - 数据门槛
  - 特征门槛
  - 算法门槛
- 模型应用门槛
  - 适应性门槛
  - 信任门槛
  - 数据安全和隐私门槛

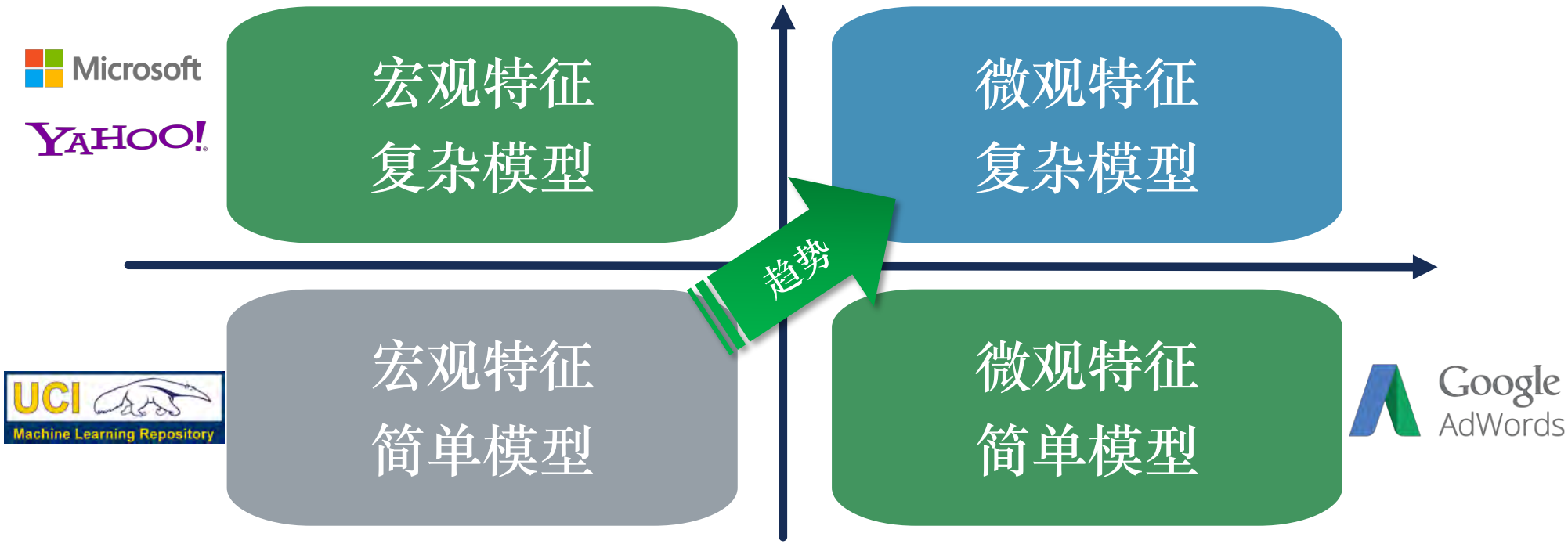
建模门槛：实际应用中数据和维度的趋势

- 有效数据的增长
  - 数据量：10^4 → 10^{10}~10^{12}
- 数据维度的增长
  - 宏观维度（10^3）→ 微观维度（ 10^{10}~10^{12}）



建模门槛：机器学习模型的趋势

机器学习模型在工业应用中的四个象限



建模门槛：没有免费的午餐

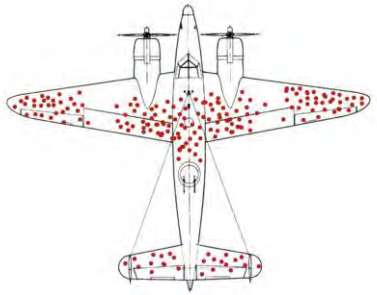
- No Free Lunch定理：[Wolpert and Macready 1997]任意两个算法a_1和a_2，

$$\sum_f P(d_m^y|f,m,a_1) = \sum_f P(d_m^y|f,m,a_2)$$

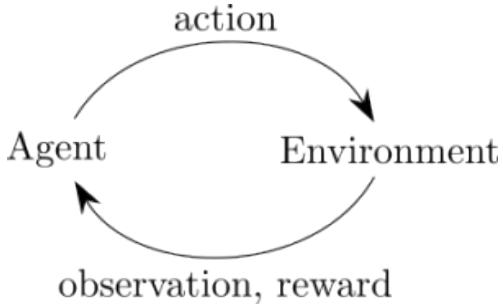
- 任意算法（包括随机算法）在**所有问题**上的期望性能一样
  - **不存在通用算法**
  - 但在具体的实际问题上，有可能存在比其他算法好的算法
    - 需要针对不同的实际问题，研究开发不同的机器学习算法

适应性门槛：面对开放世界

- 数据分布变化
  - 迁移学习
    - Importance Sampling
- 与环境交互、新训练样本
  - 强化学习
- 新训练目标
  - 迁移学习
- 样本属性含义变化



[World War II, Abraham Wald]



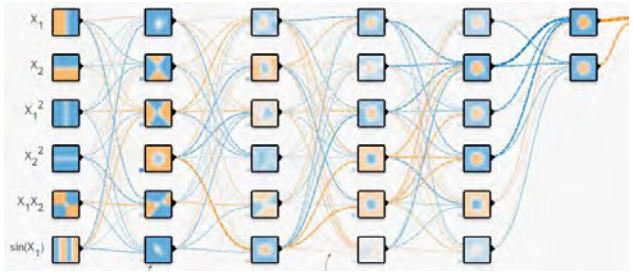
[强化学习]

降低适应性门槛：鲁棒机器学习

- 训练阶段
  - 对噪声数据的鲁棒性
- 应用阶段
  - 模型对未知样本的鲁棒性
    - 置信度估计
  - 对关键性高风险应用的鲁棒性
    - 增加数据、Safe Machine Learning算法

信任门槛：黑箱模型

- 比如医疗应用：只给出诊断，不给出原因无法给出治疗方案
- 可解释机器学习
  - Twice Learning [Zhou, 2004]
  - LIME [Ribeiro, 2016]
  - Influence Functions Interpretation[Pang Wei Koh, 2017]



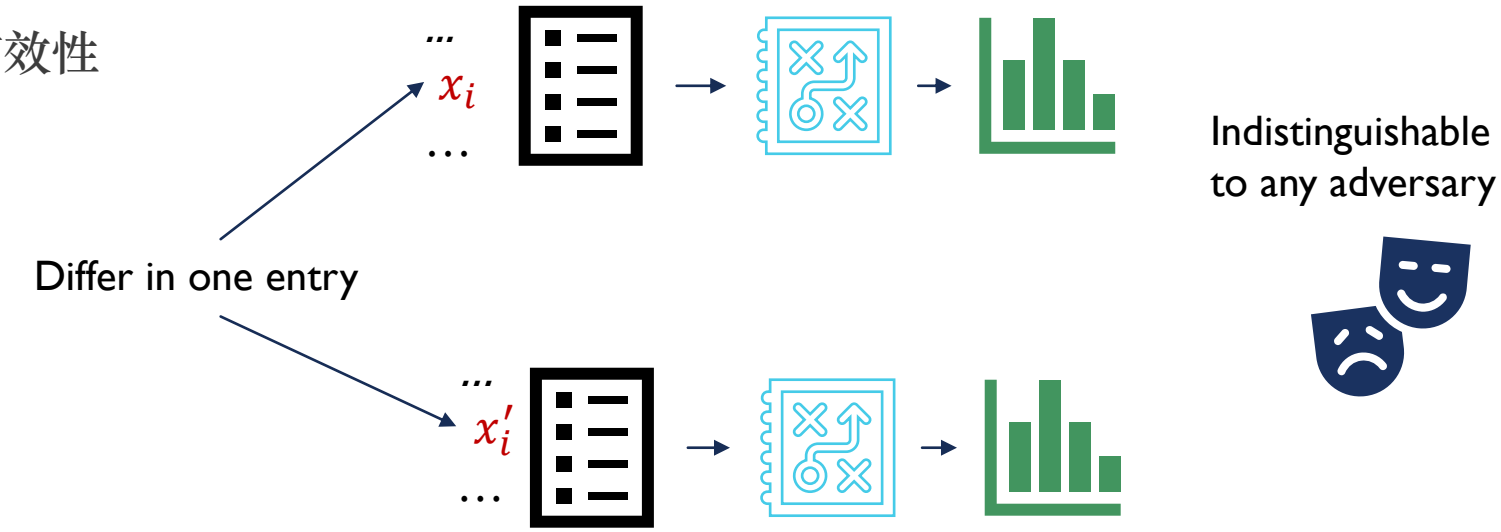
[Tensorflow DNN]

[Decision Tree]

[Zhi-Hua Zhou, 2004]

数据安全和隐私门槛

- 保护用户隐私，同时保持数据的有效性
  - 解决方案
    - 保留数据隐私的机器学习方法
      - Differential Privacy
    - 模型交易取代数据交易



[Dwork 2006]

目录

- AI应用背景介绍
- **AI for everyone**
  - 效果
  - **成本**
- 总结

机器学习应用的成本

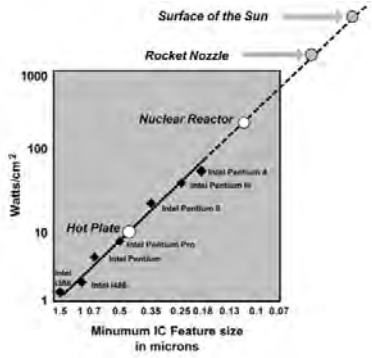
- **计算成本**
- 专家成本
  - 编程门槛
  - 机器学习专业门槛
- 数据成本

降低计算成本：计算效率优化

- **计算**
- 存储
- 通讯
- 容错

分布式并行计算

- 摩尔定律失效
  - 能耗墙（Power Wall）
  - 延迟墙（Latency Wall）
- 单机能力有限
  - IO、存储、计算有限
- 目前提升计算能力的主流方式
  - 并行化：降低执行延迟→提升吞吐
  - 但是，Amdahl定律



[Power Wall]

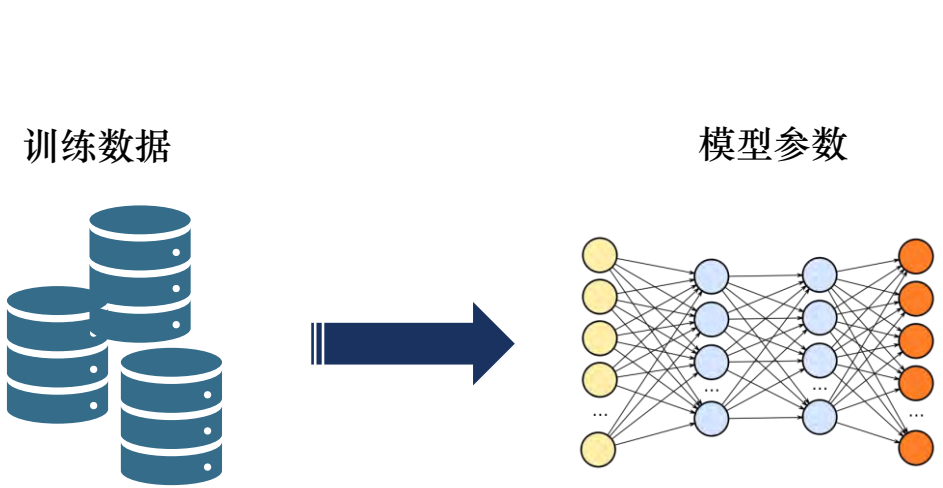
[Latency Wall]

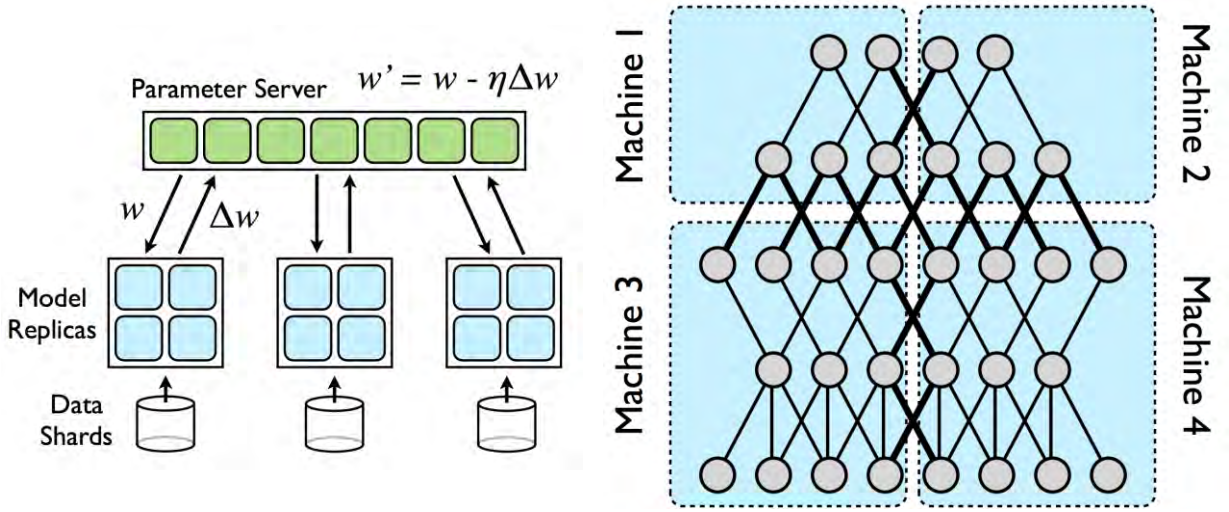
$$S_{\text{latency}}(s) = \frac{1}{(1-p) + \frac{p}{s}}$$

[Amdahl定律]

分布式并行模型训练

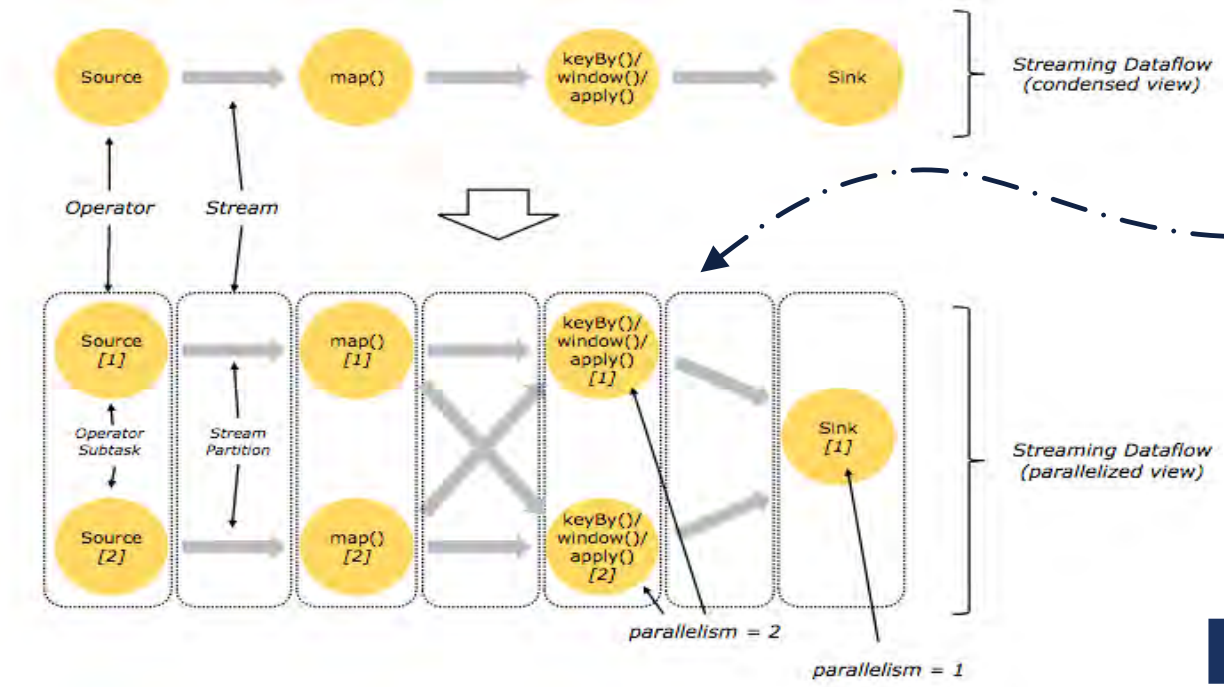
- 数据分布式和模型分布式

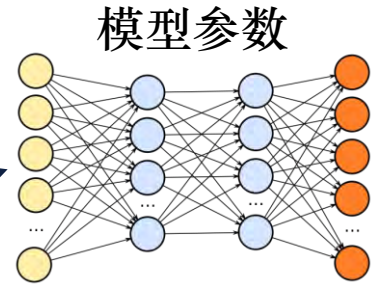




[Large Scale Distributed Deep Networks, Google]

典型计算模型：数据流





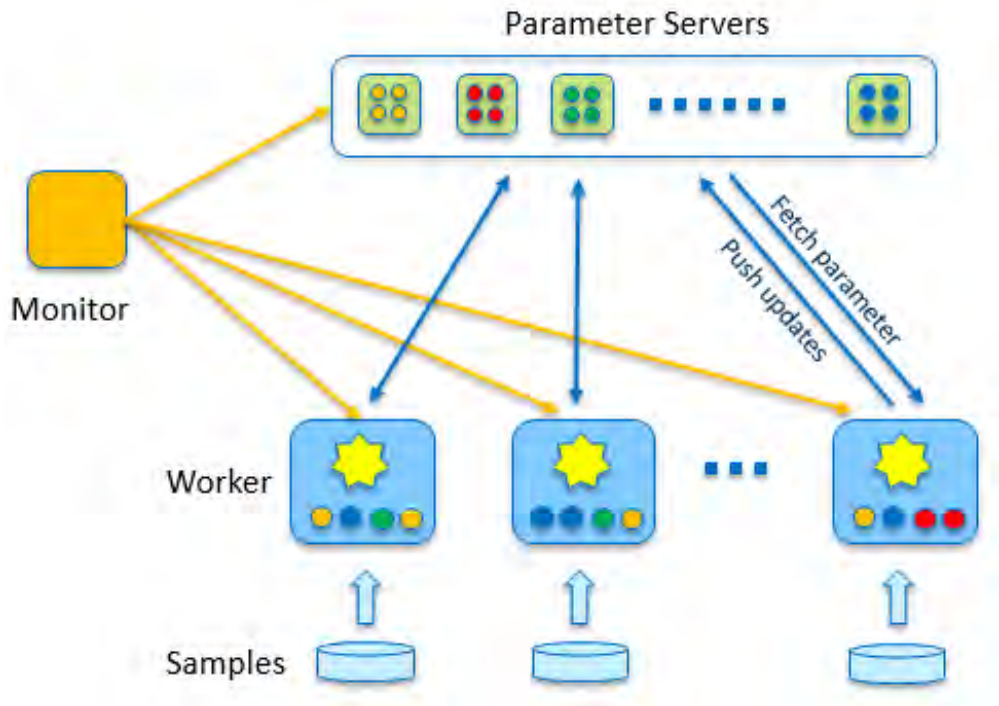
典型机器学习模型优化过程：

- 查取原模型w^{old}
- 根据原模型计算更新Δw
- 更新模型$w^{new} = w^{old} + \Delta w$

问题

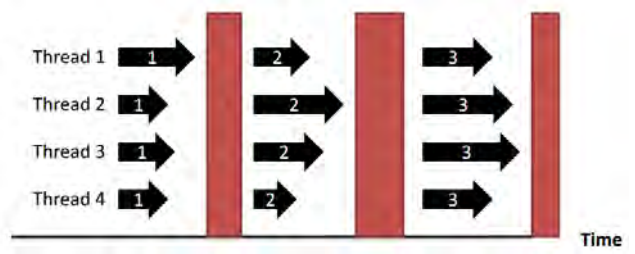
模型参数w是一个所有计算共享的中间状态

典型计算模型：参数服务器

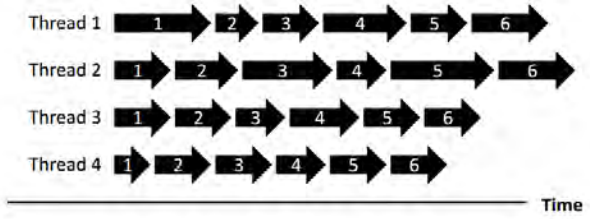


Google MLC PETUUM

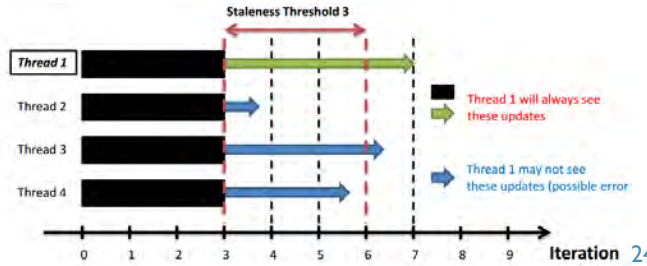
一致性模型



[BSP]

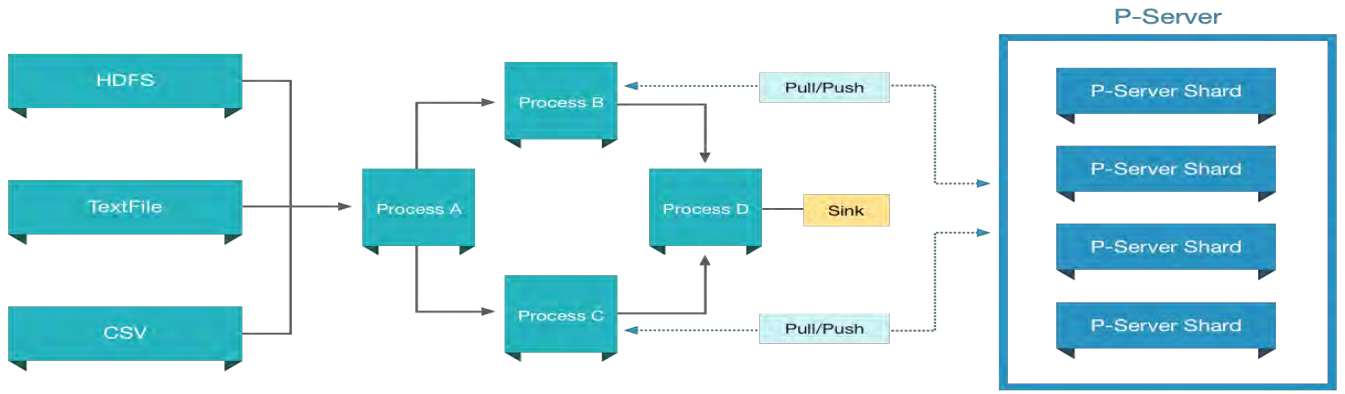


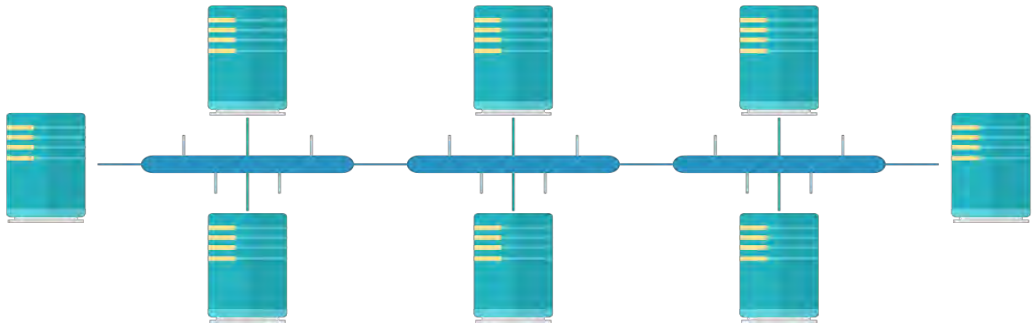
[ASP]



[SSP, Xing]

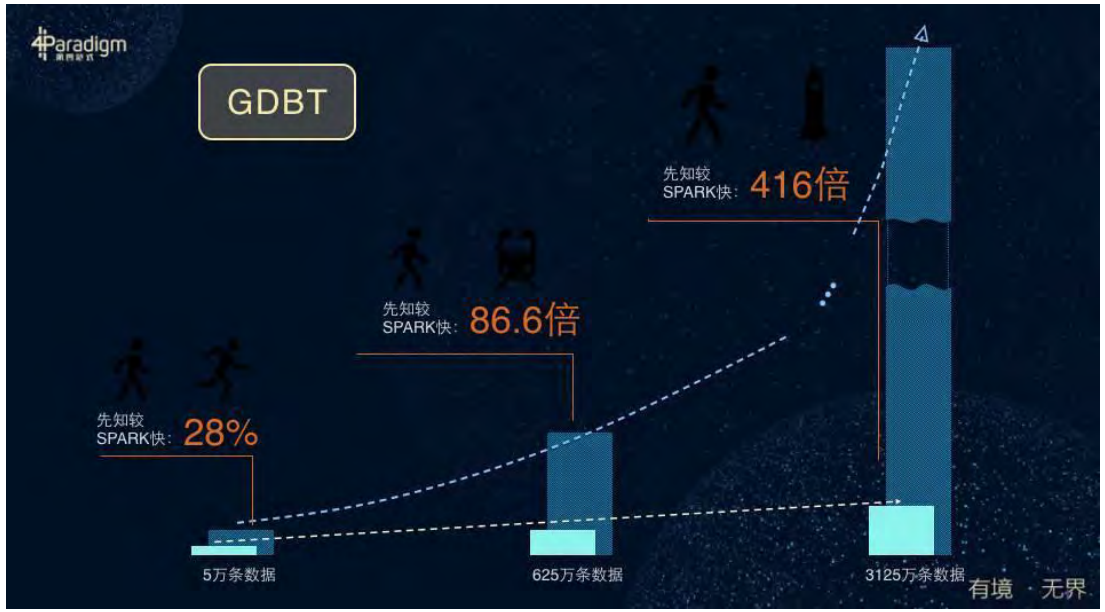
趋势：数据流 + 参数服务器


P-Server
HDFS
TextFile
CSV
Process A
Process B
Process C
Process D
Pull/Push
Pull/Push
Sink
P-Server Shard
P-Server Shard
P-Server Shard
P-Server Shard



其他计算效率优化

- 计算
  - 异构计算优化
  - 异步，合理地计算调度
- 存储
  - 不同存储设备共存：Hard Disk / SSD / NVMe / RAM / L2 Cache…
  - 多级缓存
- 通讯
  - 提升网络吞吐、降低网络延迟
    - 软件：请求合并、缓存
    - 硬件：多网卡、InfiniBand…
- 灾备
  - Data Lineage VS. Checkpointing



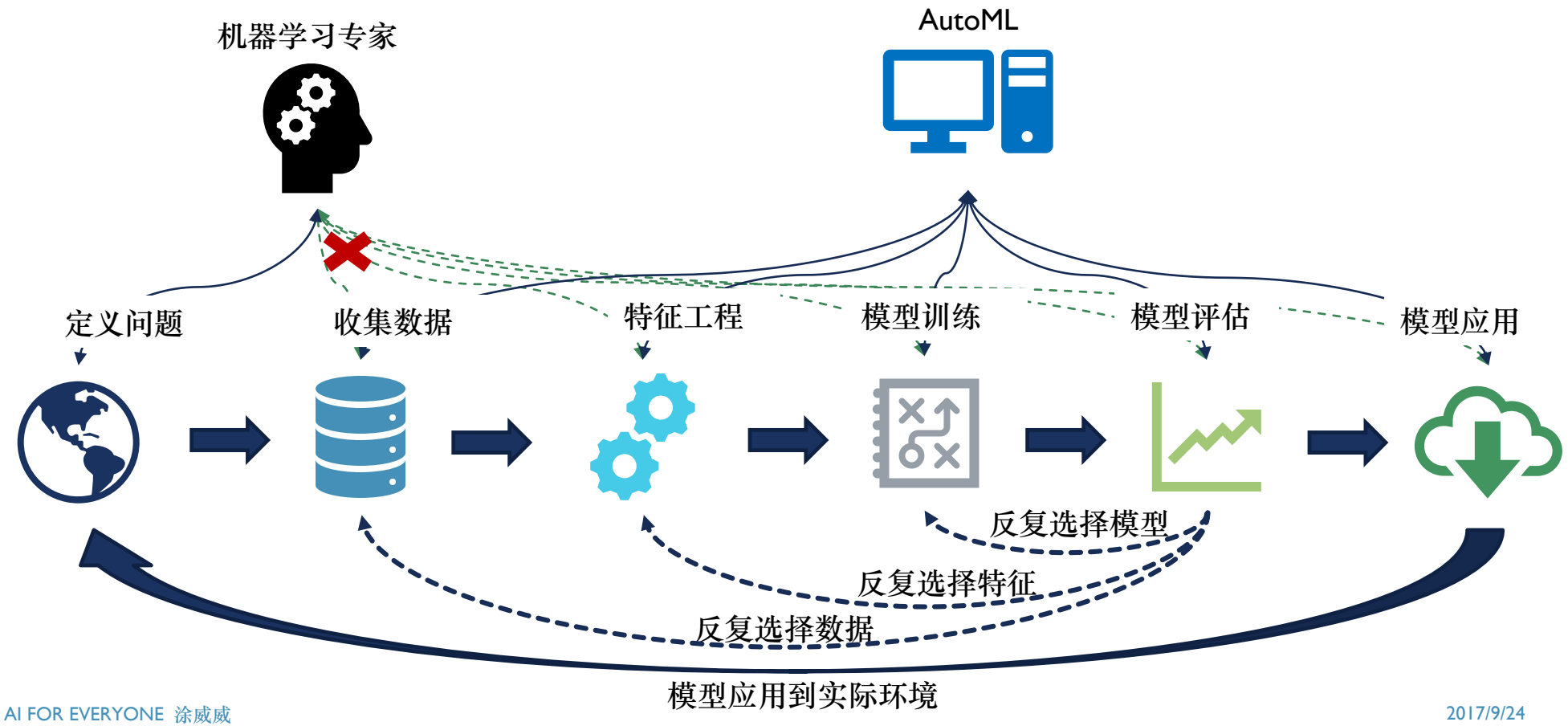
机器学习应用的成本

- 计算成本
- **专家成本**
  - 编程门槛
  - 机器学习专业门槛
- 数据成本

降低编程门槛：机器学习平台

[The Fourth Paradigm]

降低专业门槛：从“人工”智能到机器智能

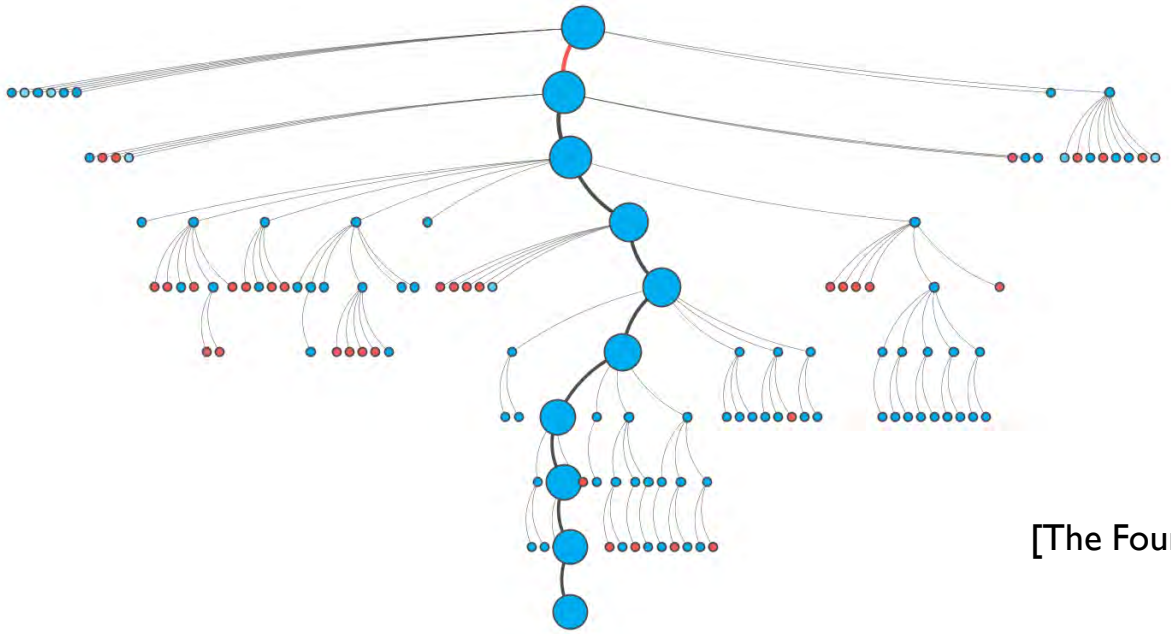


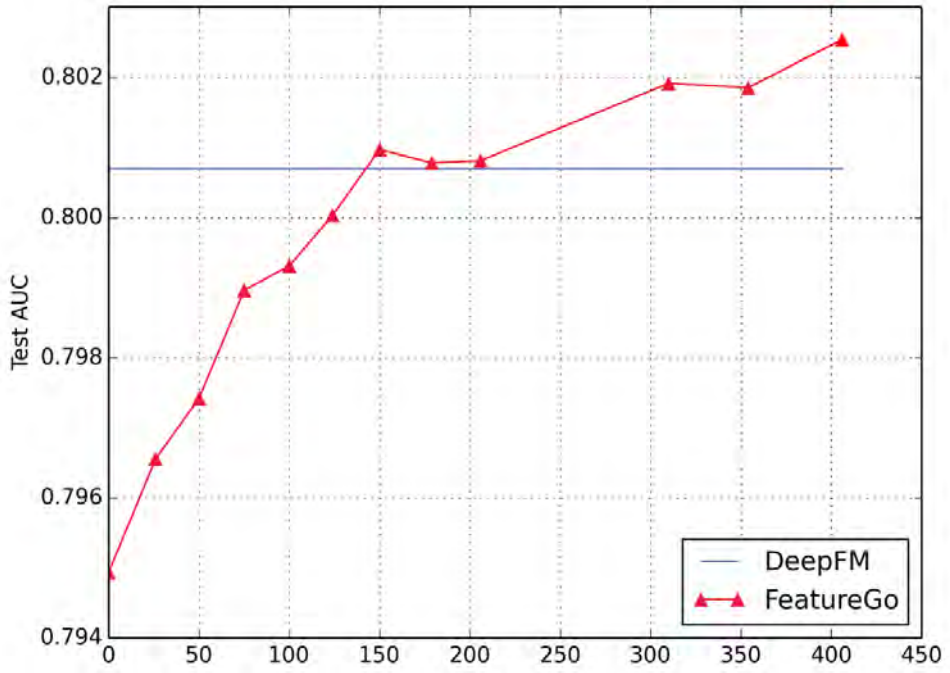
自动机器学习（AUTOML）

- 自动数据清洗
- 自动数据类型推断
- **自动特征工程**
- **自动模型和参数选择**
- 自学习

自动组合特征

- 自动化特征组合：FeatureGo
  - 问题空间2^{2^d}
    - $d = 20, 10^{315652}$
    - AlphaGo空间10^{171}

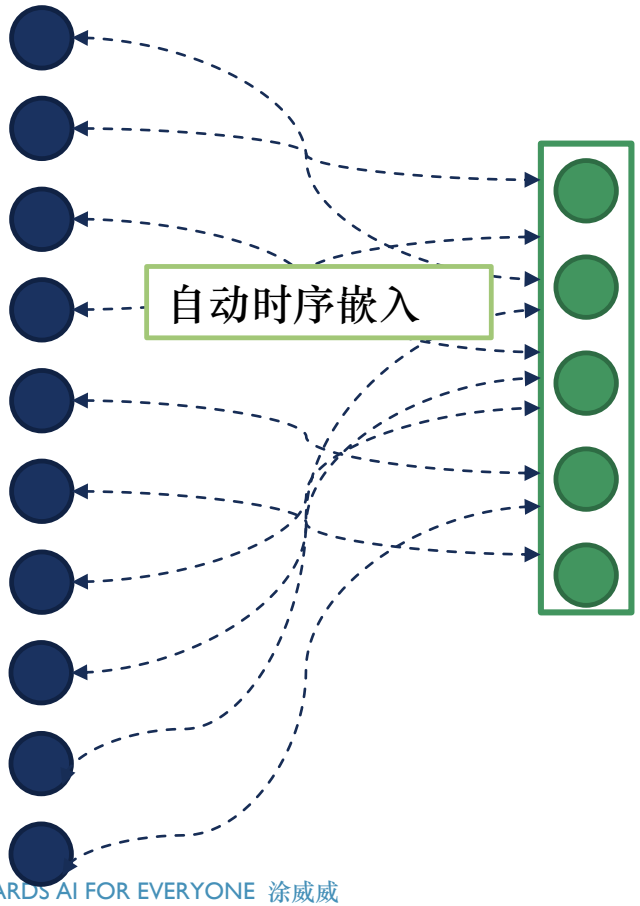


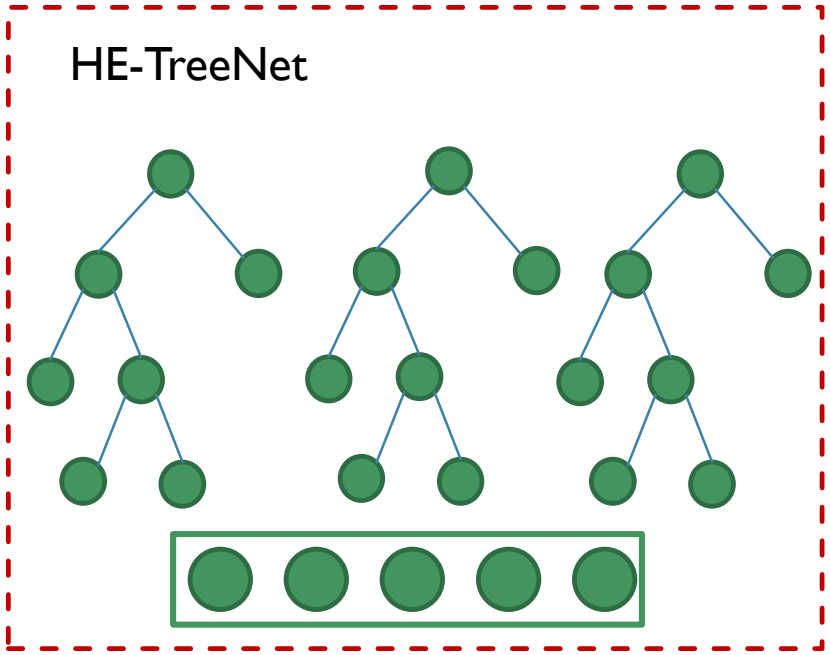


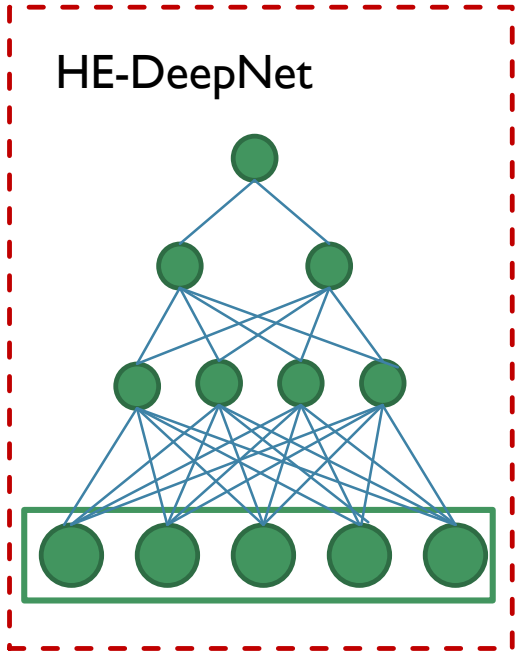
FeatureGo VS. IJCAI2017

[The Fourth Paradigm]

自动时序特征





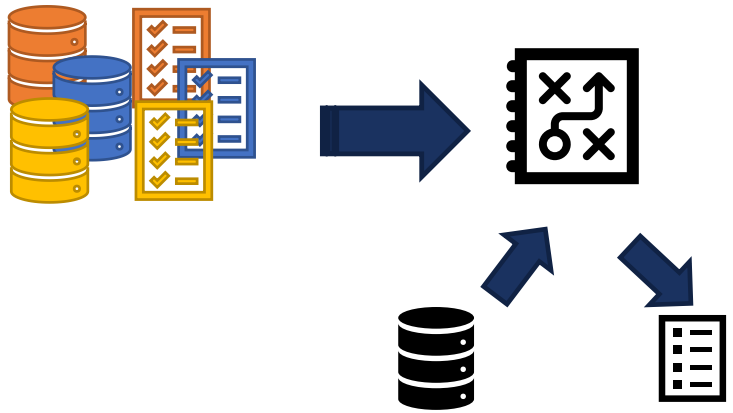


[The Fourth Paradigm]

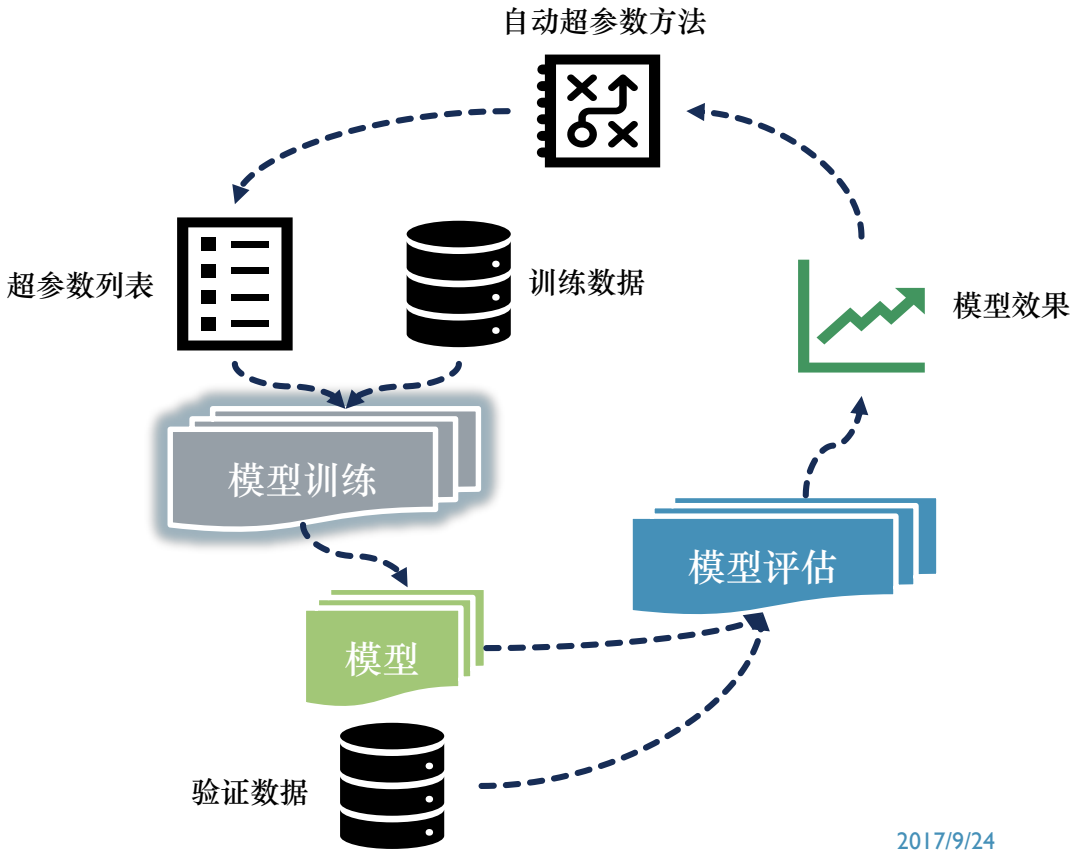
X AUC improvement: ~2%

自动模型和超参数选择

- Bayes方法
- 演化计算方法
- 迁移学习方法

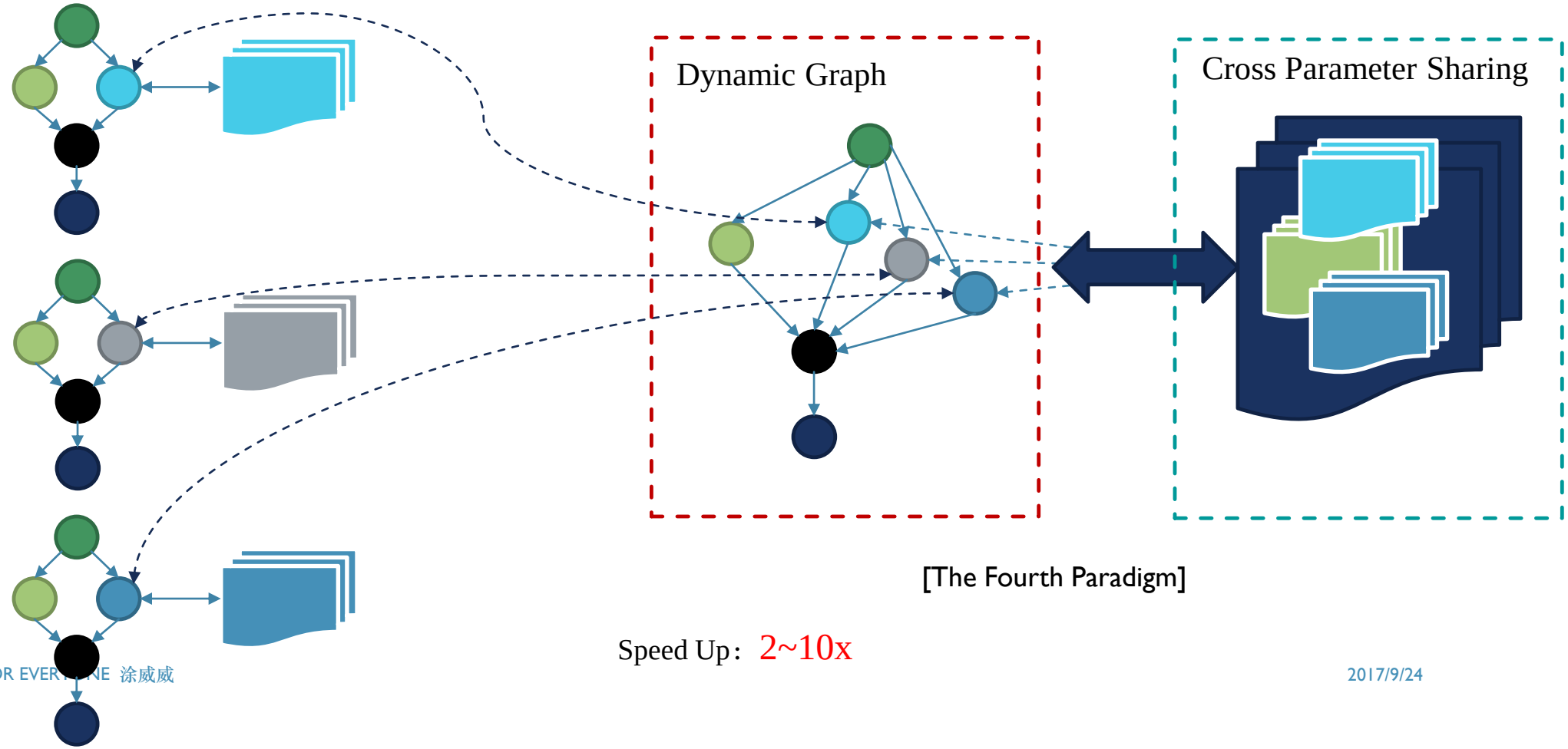


[迁移学习超参数选择方法]



[Bayes、演化计算超参数选择方法]

自动模型和参数选择：工程优化



[The Fourth Paradigm]

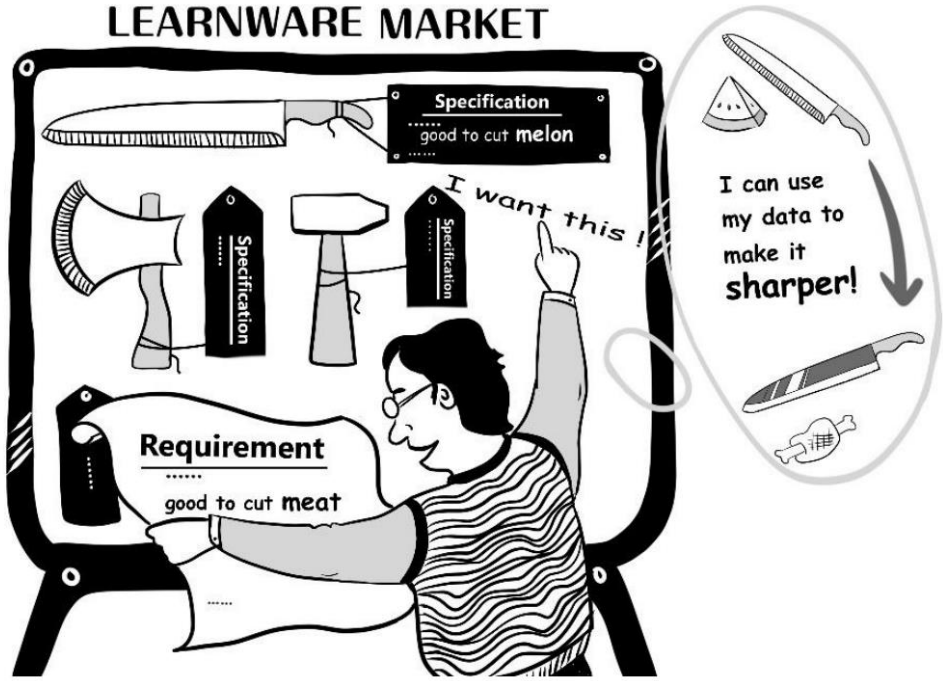
Speed Up：2~10x

机器学习应用的成本

- 计算成本
- 专家成本
  - 编程门槛
  - 机器学习专业门槛
- **数据成本**

降低数据成本：学件、迁移学习

- 学件 = 模型（Model）+ 规约（Specification）
  - 可重用
  - 可演进
  - 可了解
- 迁移学习[Pan & Yang, TKDE 2010]
  - 特征迁移
  - 样本迁移
  - 强化迁移学习
  - 终生学习



[Zhi-Hua Zhou, FCS 2016]

总结

- AI在工业界有了很多成功的应用
- AI for Everyone
  - 效果
    - 高维复杂模型
    - 强化学习
    - 鲁棒机器学习
    - 可解释机器学习
  - 成本
    - 降低专家成本：AutoML
    - 降低计算成本：计算效率优化
    - 降低数据成本：学件、迁移学习

中生代技术
FRESHMAN TECHNOLOGY

ArchData技术峰会全国巡回
上海9月，北京9月，成都10月，南京10月，
长沙11月，广州11月
中生代咨询内训
技术架构，研发管理，敏捷开发，大数据
微服务，AI，机器学习
中生代人才内推
对接研发主管，内推精准人才