小而美的机器学习平台实践

谭孟泷

关于我

- 转转算法部负责人
- 前人人车业务平台技术总监
- 前阿里妈妈移动广告算法专家
- 前百度移动搜索资深策略工程师
- 毕业于武汉大学

分享内容

1. 背景介绍
2. 算法平台化尝试
3. 经验总结

关于转转

关于转转算法部

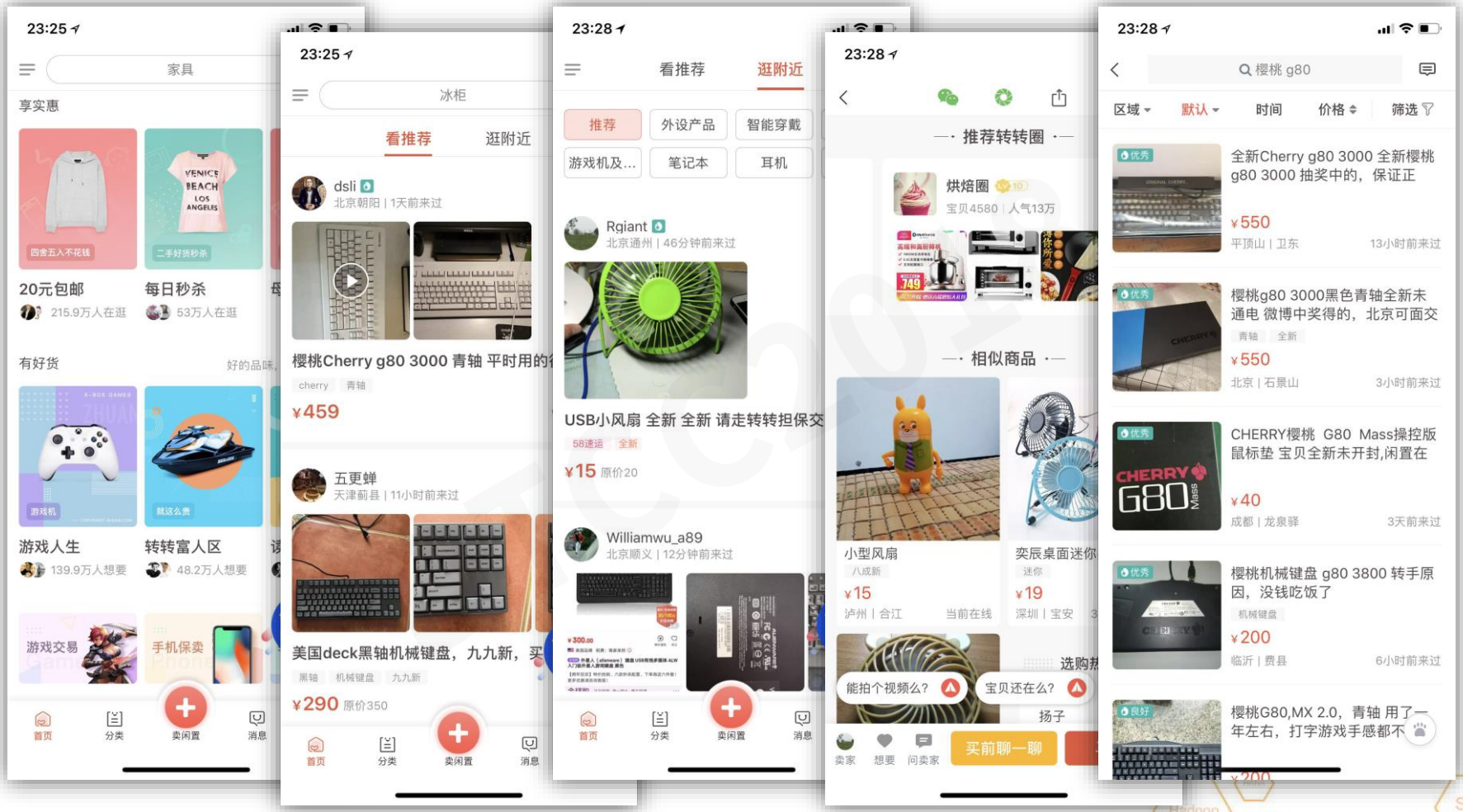
曾经：

- 按业务线切分团队
- 纯业务收益导向，追求短期业绩产出
- 全栈机器学习工程师

现在：

- 统一算法部门，横向支持多条业务线
- 兼顾 短期业务效果 和 中长期算法中台能力建设
- 专业化分工，算法RD 和 工程RD

算法场景举例



算法场景举例，and more

- 广告投放：创意优化、受众扩散
- 展现引导：标题摘要改写、商品卖点挖掘、联想词、Feed词
- 运营活动：人群定向、智能组货、自动化红包
- 平台风控：反黄、反欺诈、违禁品识别、团伙挖掘、羊毛党识别、模拟器识别
- 客服系统：智能分单、客服机器人
- 垂直业务：辅助定价、销量预估
- ……

算法团队面临的挑战

- 需求多样且复杂，涉及算法模型面广
- 支持方式多样，既打主力，又打辅助
- 业务需求旺盛，算法研发狼多肉少

怎么破？

提高效率

- 加速算法迭代：开发、上线、调试
- 提高复用水平，减少重复工作

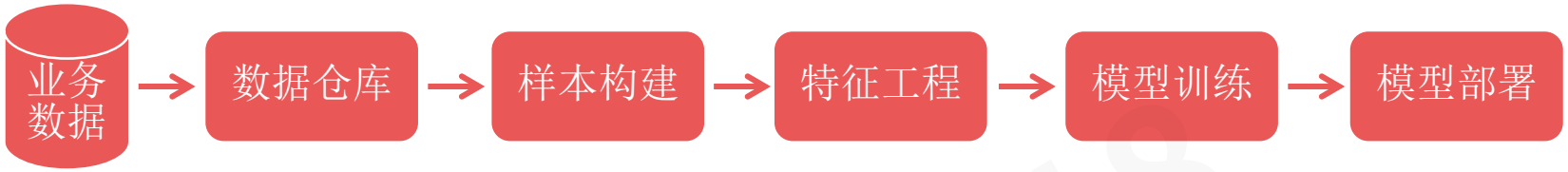
增加人手

- 扩充算法团队规模
- 算法能力中台化，赋能其他角色

Part 2

算法平台化尝试

算法迭代的效率提升点



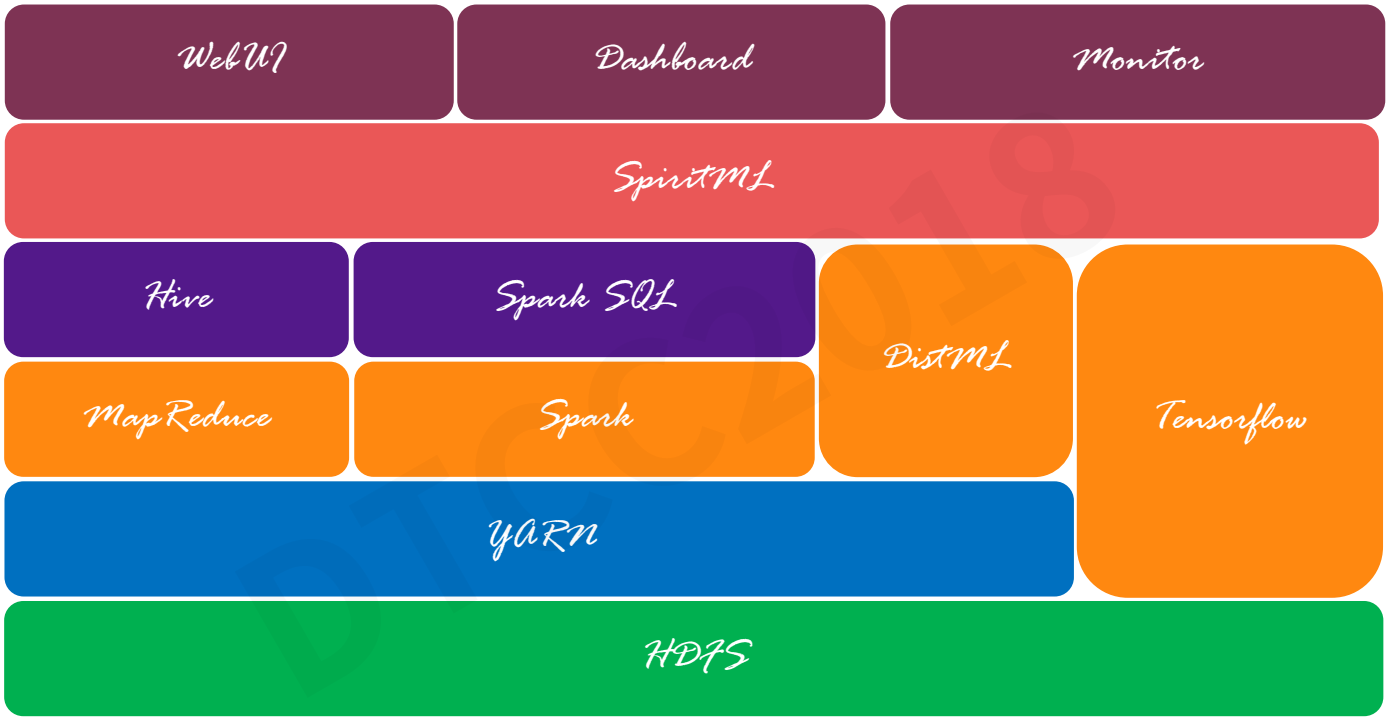
数据层面

- 统一口径，减少冗余
- 优化流程，提升性能
- 加强数据治理，减少沟通成本

算法流程开发

- 组件化流程化，提升复用水品
- 简化编程接口，降低使用门槛
- 抽象常见业务场景的通用解决方案
- 模型部署双透明：算法细节对工程同学透明，线上服务细节对算法同学透明

机器学习工具链



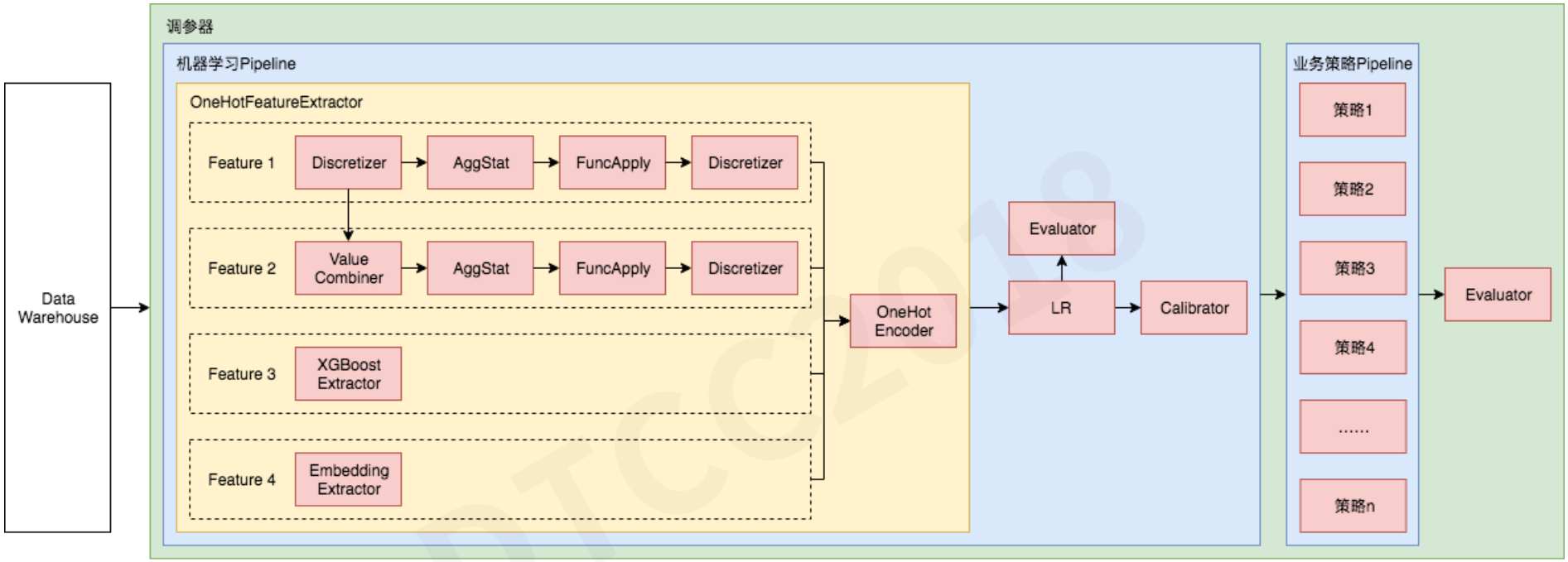
SpiritML

- 提供MLlib风格 和 Hive风格接口
- 以算子为基本单元，支持DAG
- 封装常用算法功能，一站式覆盖常见算法场景
- 以DataFrame为数据操作接口，封装IO细节
- 提供serving能力，实现模型部署 “零 ”代码

算子系统

业务算子	CTR预估	文本分类	图像分类	异常检测	.....
组合算子	样本构建	特征工程	调参	流程优化	.....
基础算子	抽样	特征处理	模型训练	效果评估	.....
	数据转换	统计操作	数值优化	序列化	.....
底层算子	map	flatMap	union	groupBy	.....
	MatMul	Conv2D	SoftMax	Argmax	.....

开发流程

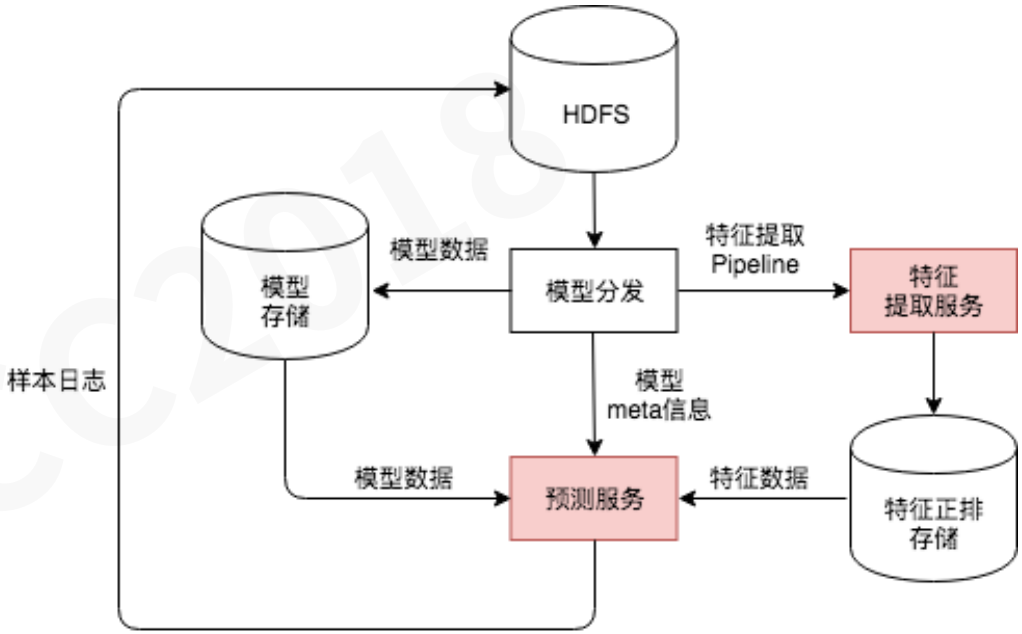


1. 通过算子串联和组合实现功能开发
2. 机器学习 和 业务策略 统一的*pipeline*
3. 参数空间抽象 和 全流程调参

模型分发

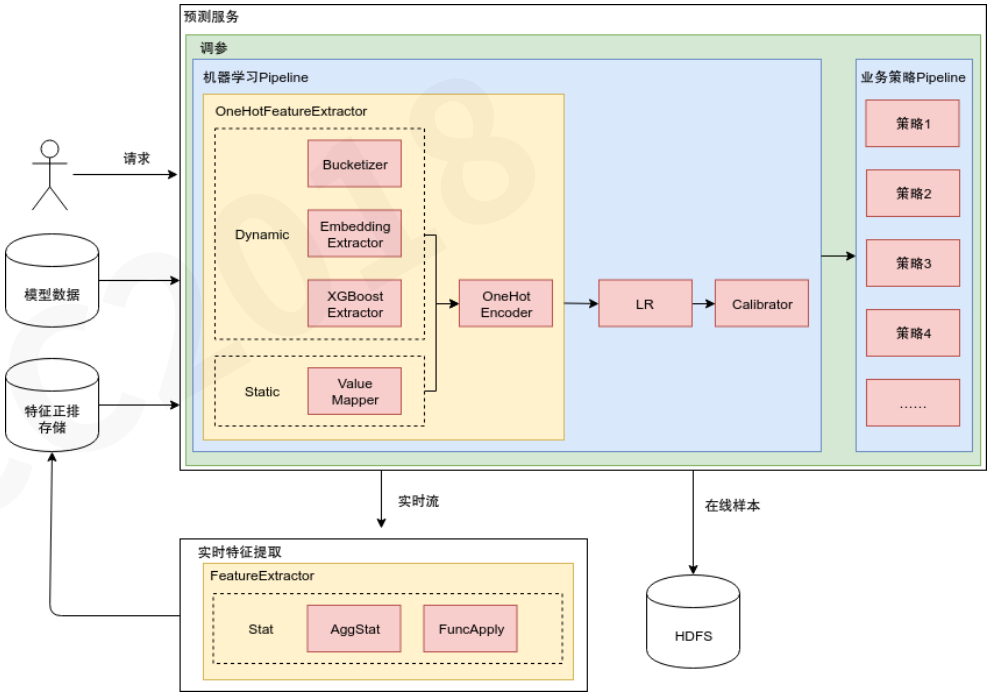
需要分发的3个部分：

1. 特征提取Pipeline
2. 预测Pipeline
3. 模型meta 和 模型数据



模型服务

- 预测*Pipeline*
  1. 动态 和 静态 流程
  2. 本地版*DataFrame*，融合特征正排服务的*API*
  3. 性能优化：特征*cache*、简单模型和流程整合、向量化
- 实时特征提取*Pipeline*
- 特征正排存储
- 在线样本



Part 3

经验总结

在初创团队做算法工程的一些心得

1. 算法和工程尽量同时起步
2. 算法的不确定性和工程的确定性
3. 工程部分要独立于算法进行测试
4. 追求算法流程代码的极致复用
5. 工程部分要扮演好监督者角色
6. 算法流程中要刻意去发现并设置*checkpoint*
7. 尽量减小训练数据和实际预测之间的*gap*
8. 要多关爱做工程的同学

THANKS


HANA
Hadoop
SQL
MySQL
DB2
BigData
Spark
SQL Server
NewSQL
Oracle
MPP
NoSQL
OLAP

讲师申请
联系电话（微信号）：18612470168

关注“ITPUB”更多
技术干货等你来拿~

与百度外卖、京东、魅族等先后合作系列分享活动

微学堂

让学习更简单

微学堂是以ChinaUnix、ITPUB所组建的微信群为载体，定期邀请嘉宾对热点话题、技术难题、新产品发布等进行移动端的在线直播活动。

截至目前，累计举办活动期数60+，参与人次40000+。

ITPUB学院

ITPUB学院是盛拓传媒IT168企业事业部（ITPUB）旗下
企业级在线学习咨询平台
历经18年技术社区平台发展
汇聚5000万技术用户
紧随企业一线IT技术需求
打造全方式技术培训与技术咨询服务
提供包括企业应用方案培训咨询（包括企业内训）
个人实战技能培训（包括认证培训）
在内的全方位IT技术培训咨询服务

ITPUB学院讲师均来自于企业
一些工程师、架构师、技术经理和CTO
大会演讲专家1800+
社区版主和博客专家500+

培训特色

无限次免费播放
随时随地在线观看
碎片化时间集中学习
聚焦知识点详细解读
讲师在线答疑
强大的技术人脉圈

八大课程体系

基础架构设计与建设
大数据平台
应用架构设计与开发
系统运维与数据库
传统企业数字化转型
人工智能
区块链
移动开发与SEO

联系我们

联系人：黄老师
电　话：010-59127187
邮　箱：edu@itpub.net
网　址：edu.itpub.net
培训微信号：18500940168