风控场景的模型平台架构

胡四海
阿里巴巴 安全部

北京 伦敦 纽约 旧金山 圣保罗 上海 东京

QCon

全球软件开发大会

[上海站]

信息安全

机器学习

人工智能

黑产

互联网金融（FinTech）

团队管理

基础设施

云计算

软件性能

硅谷

微服务

互联网架构

主办方 Geekbang. 极客邦科技 InfoQ

2017年10月17-19日
上海·宝华万豪酒店

扫描二维码
开启软件开发新思路


EGO聚变
TOGETHER CHANGE
2017
Geekbang 极客邦科技
EGO EXTRA GEEKS' ORGANIZATION NETWORKS
EGO会员招募季
EGO旨在组建全球最具影响力的技术领导者社交网络，联结杰出的技术领导者学习和成长。
2017年6月30-7月10
扫码报名

SPEAKER
INTRODUCE

胡四海

- 2010 年加入阿里，现负责阿里巴巴集团安全部风控引擎、人机对抗技术。
- 在阿里就职期间，一直从事风控相关领域的研究与风控引擎的开发，建造阿里多项核心风控产品。
- 在风控领域拥有 10 项技术发明专利，遍布风险防控各个领域。

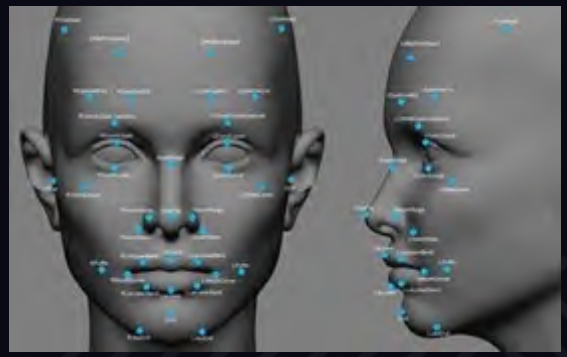
风控场景

红包雨 羊毛党

反欺诈

盗号

人机识别



活体识别

禁限售

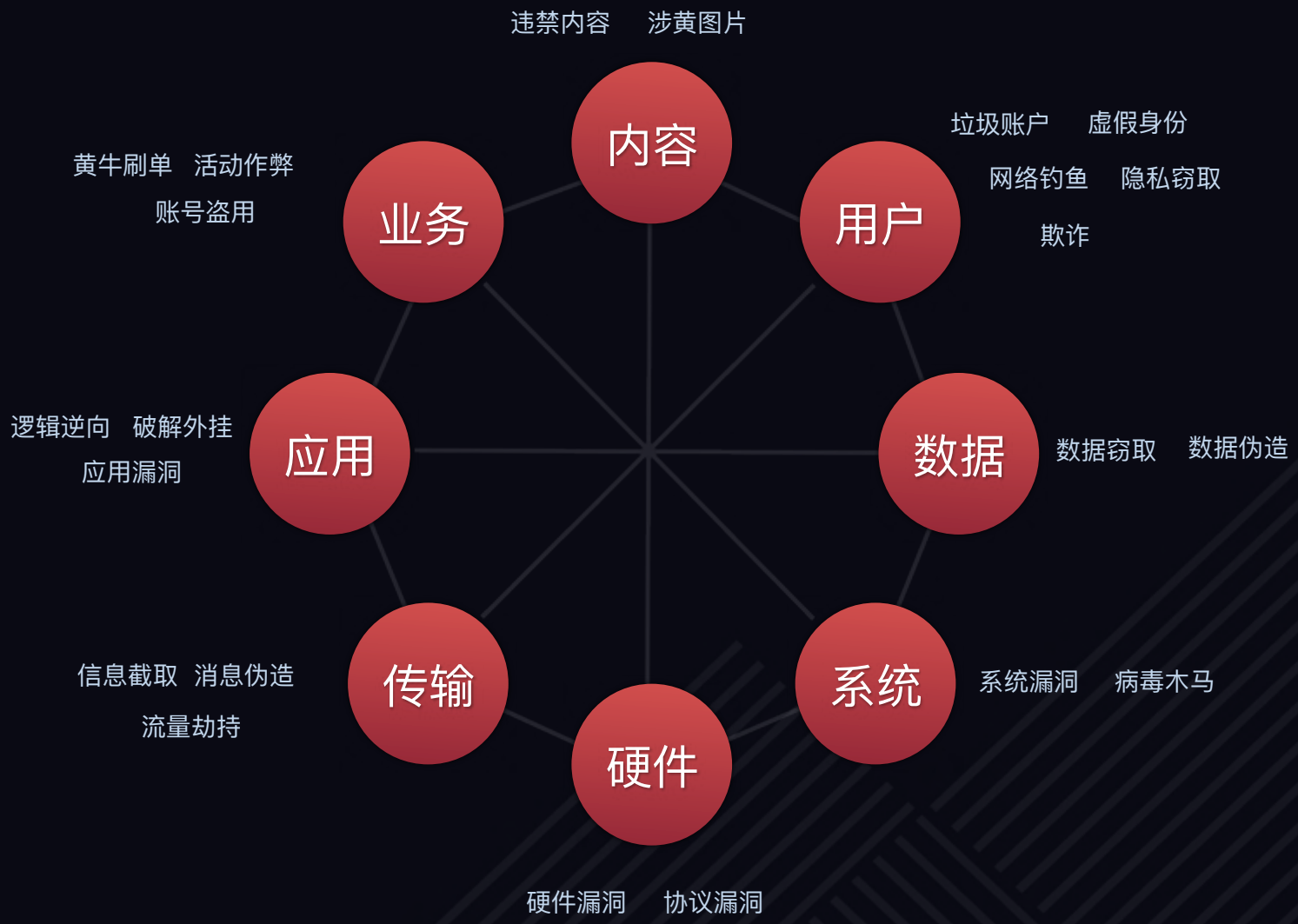
TABLE OF
CONTENTS 大纲

传统安全
是以系统为中心

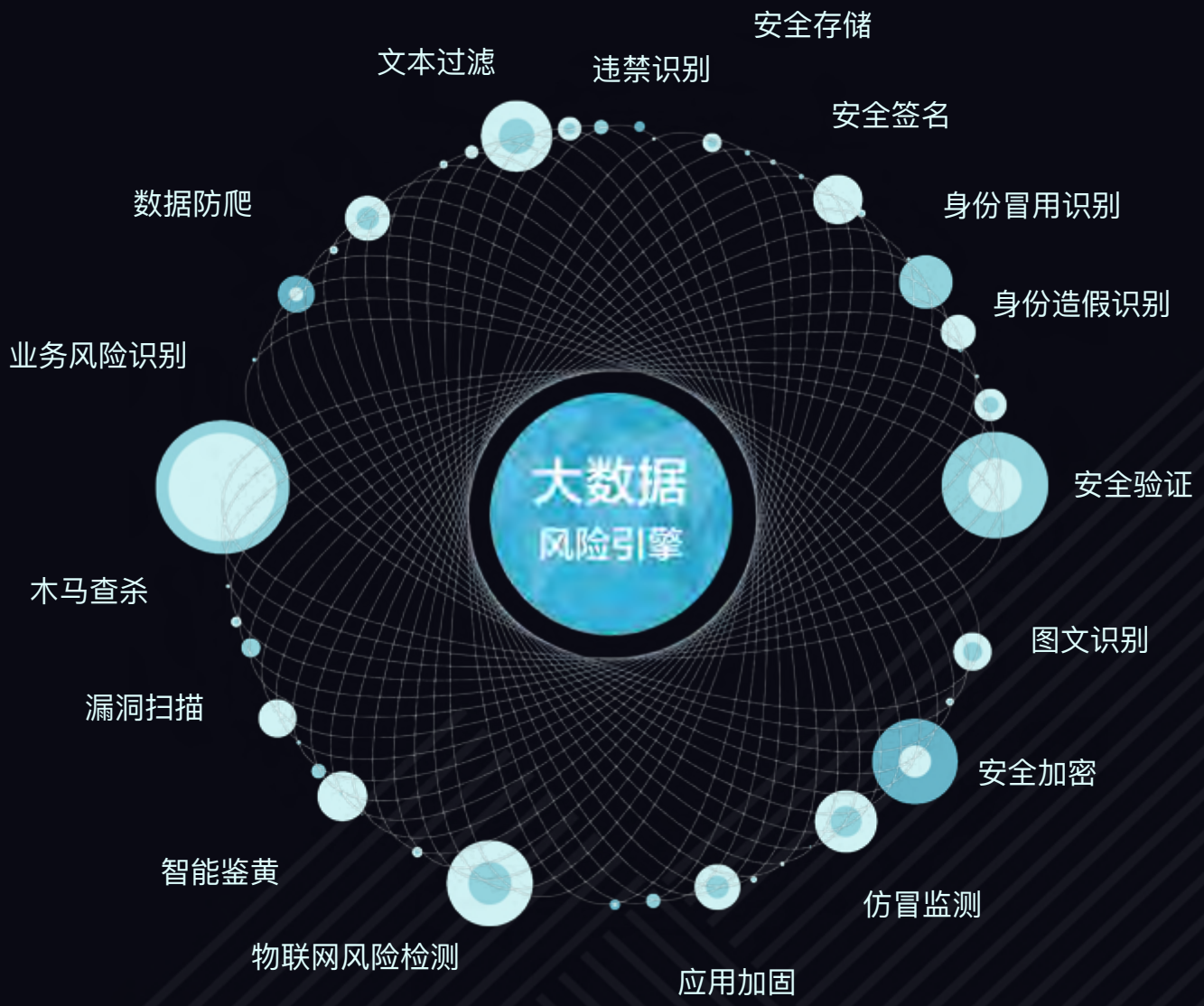
互联网时代

全互联网已泄漏个人账号超过20亿条
覆盖全互联网账号的40%以上

DT时代的安全挑战


违禁内容
涉黄图片
内容
垃圾账户
虚假身份
网络钓鱼
隐私窃取
欺诈
用户
黄牛刷单
活动作弊
账号盗用
业务
逻辑逆向
破解外挂
应用漏洞
应用
数据
数据窃取
数据伪造
信息截取
消息伪造
流量劫持
传输
系统
系统漏洞
病毒木马
硬件
硬件漏洞
协议漏洞

全链路防护体系


安全存储
文本过滤
违禁识别
安全签名
数据防爬
身份冒用识别
身份造假识别
业务风险识别
大数据
风险引擎
安全验证
木马查杀
图文识别
漏洞扫描
安全加密
智能鉴黄
仿冒监测
物联网风险检测
应用加固

端到端的风险防控

攻防布局

事前

通过积累的黑产数据，在行为发生前直接屏蔽。
发违规产品、发广告贴前进行限制。限制黑产团伙领取红包权限。
等等 …

事中

用户登陆时检测是否帐号被盗。下单时检测是否存在欺诈风险。
订单评论时检测是否是垃圾广告。是否羊毛党领取红包。
等等 …

事后

产品发布上线后进行离线扫描排查，离线模型全量扫描欺诈会员。
羊毛党红包套现。
等等 …

全链路

TABLE OF
CONTENTS 大纲

规则

模型

特征

设备指纹、实时计算、知识图谱、关系网络、NLP、文本关键词 ...

规则

可解释
应急能力强
可用于快速止血

优点

缺点

依赖人的经验
管理维护麻烦
整体效果依赖单维特征的贡献

模型

处理更大量数据
抗衰变
减少人的情感介入

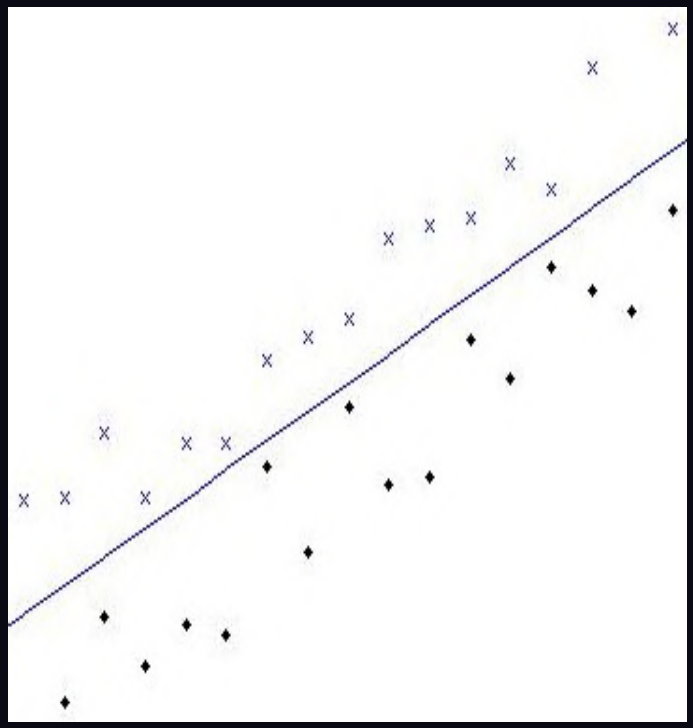
优点

缺点

可解释性弱
迭代更新周期长
知识成本高

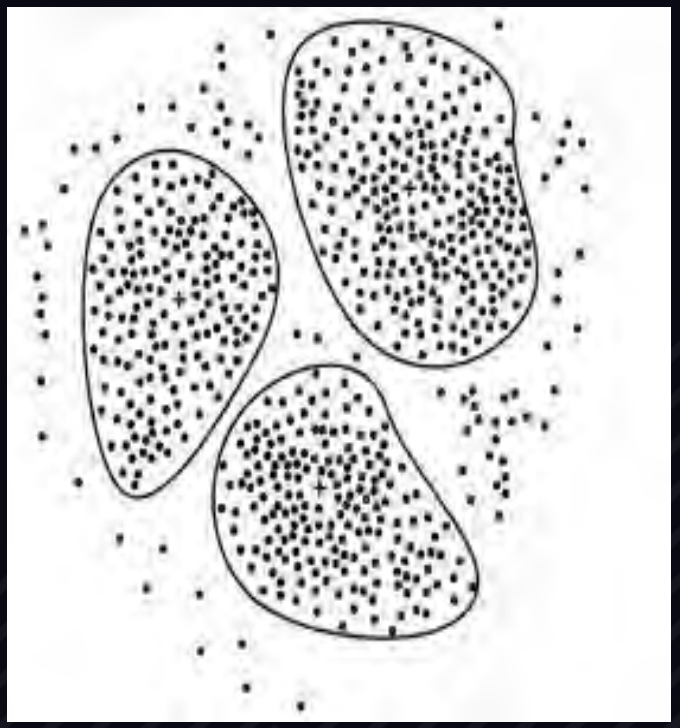
模型基本要素

有监督学习



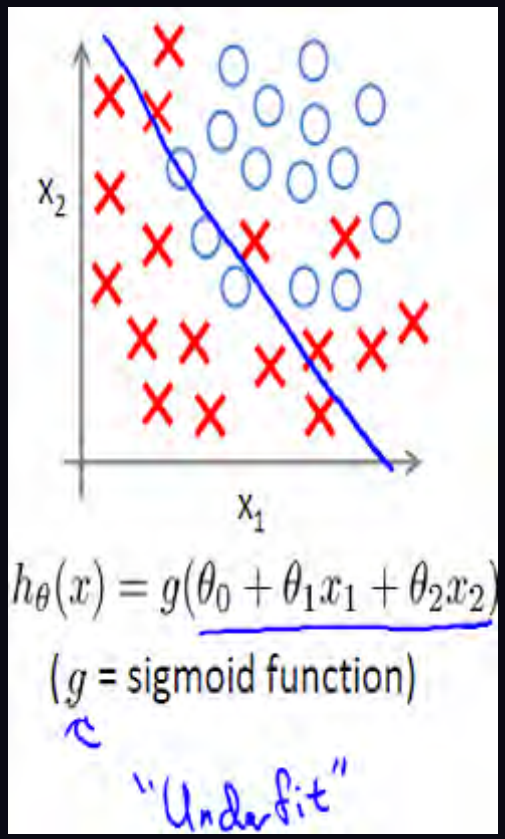
每个样本都有明确的类别

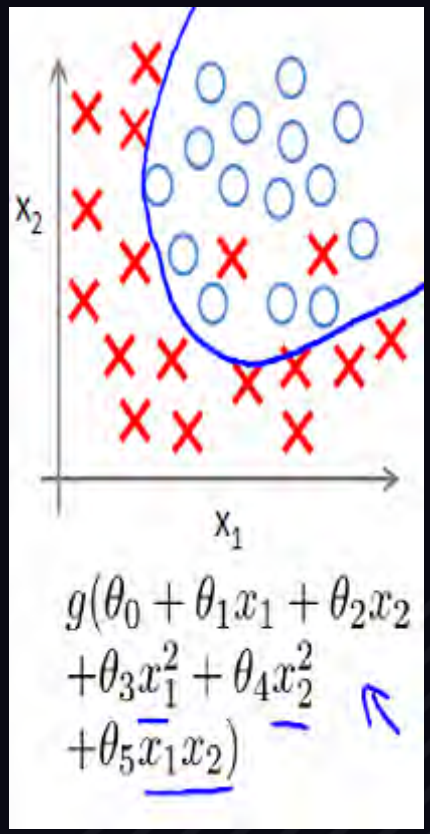
无监督学习



依靠样本间的关系划分类别

过拟合和欠拟合


X2
X1
$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$
(g = sigmoid function)
"Underfit"


X2
X1
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$

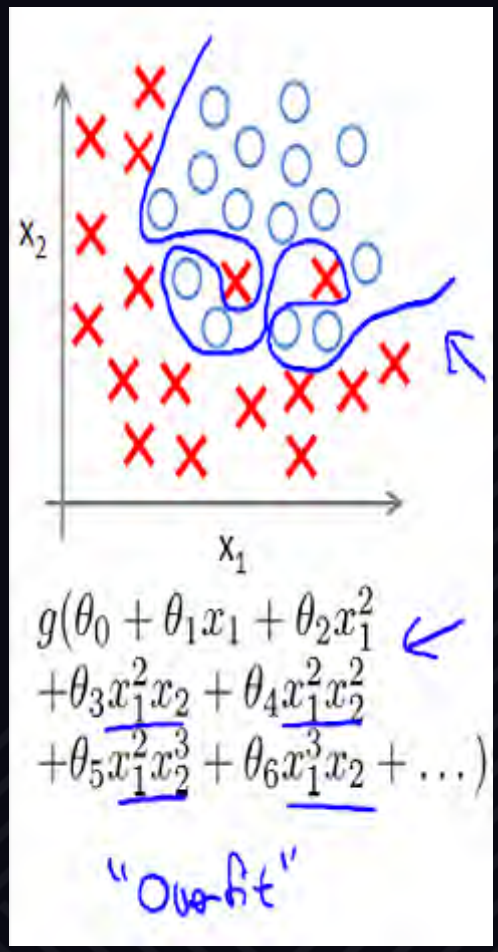

X2
X1
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \ldots)$
"Overfit"

TABLE OF
CONTENTS 大纲

模型平台定位

Incremental learning

快速上下线模型

Online learning

保证线下线上逻辑一致性

让建模人员更多关注模型防控效果本身

有效监控模型衰退情况

自动更新模型

兼容多种学习器

高效运行

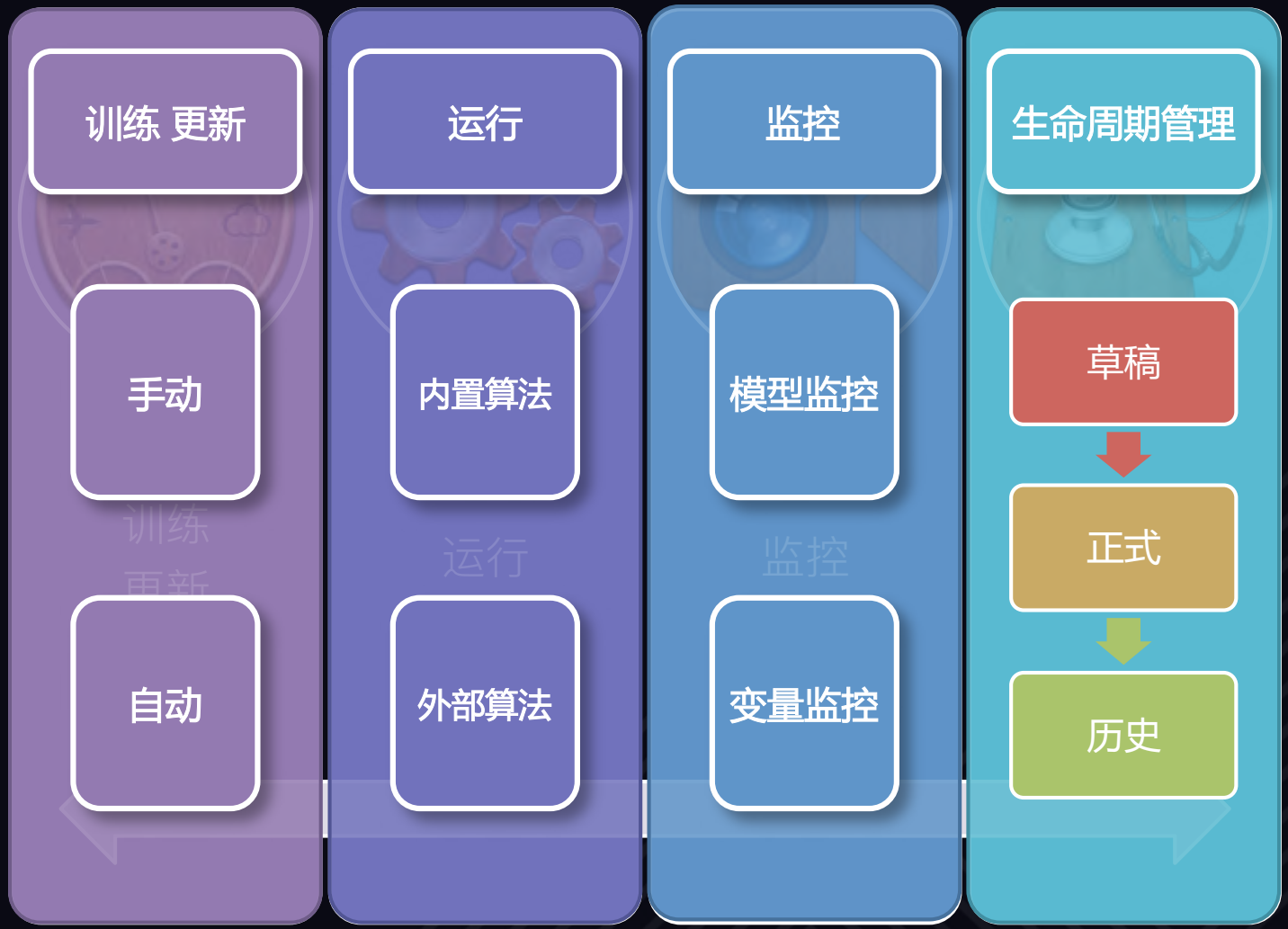
快速分流比对验证

组合模型

模型平台挑战

1. 如何简单方便地进行模型训练，新特征如何快速预热?
2. 如何进行分布式训练?
3. 训练的模型与特征如何快速上线？并保证线上线下逻辑一致?
4. 如何快速将原始数据加工成为模型所需要的数据? SQL
5. 如何做到模型可外化、可解释?
6. 运行时如何保证模型之间互不影响?
7. 如何有效地监控模型运行的情况，并在模型衰退之前作出反应?

模型平台架构



模型平台架构

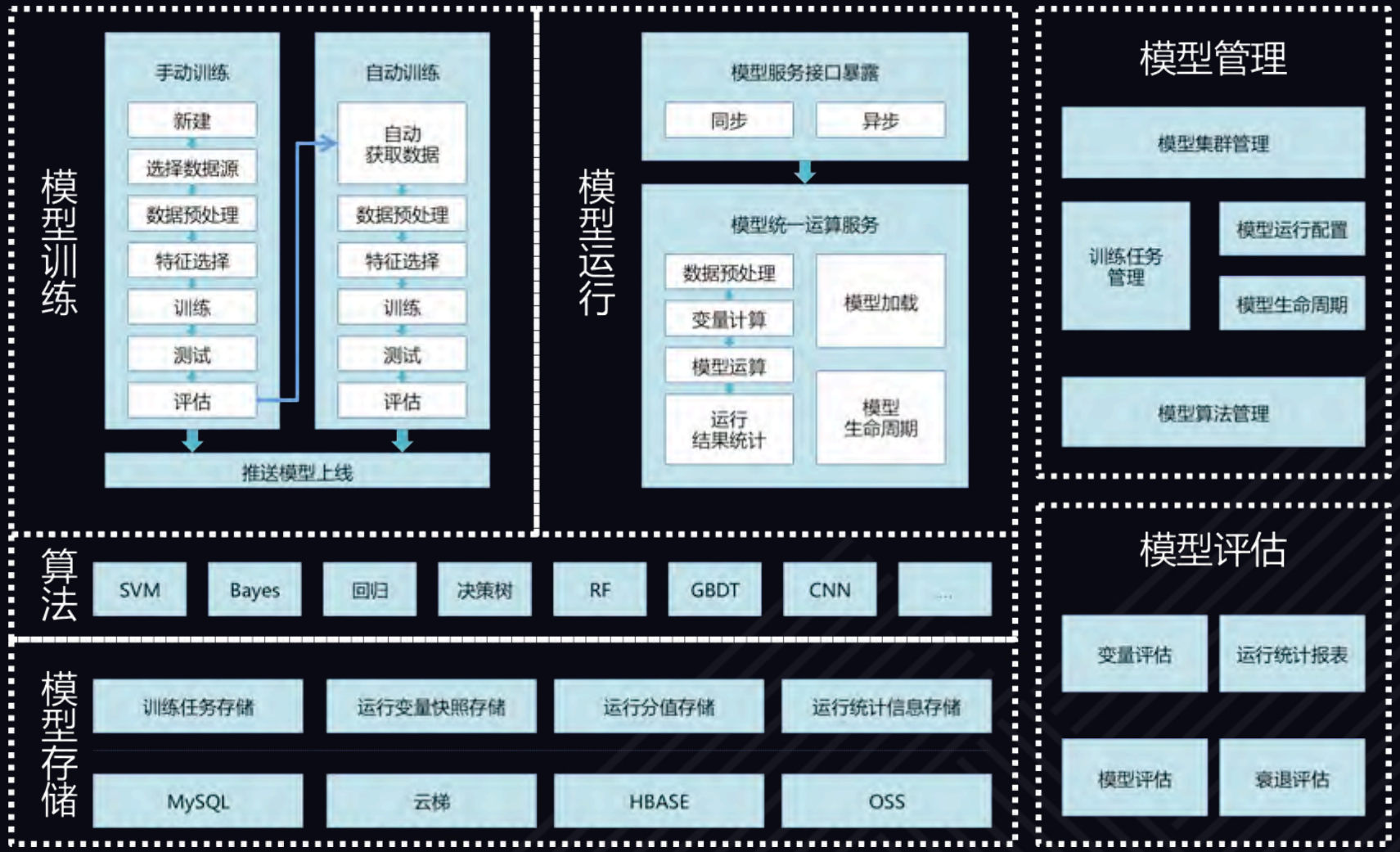
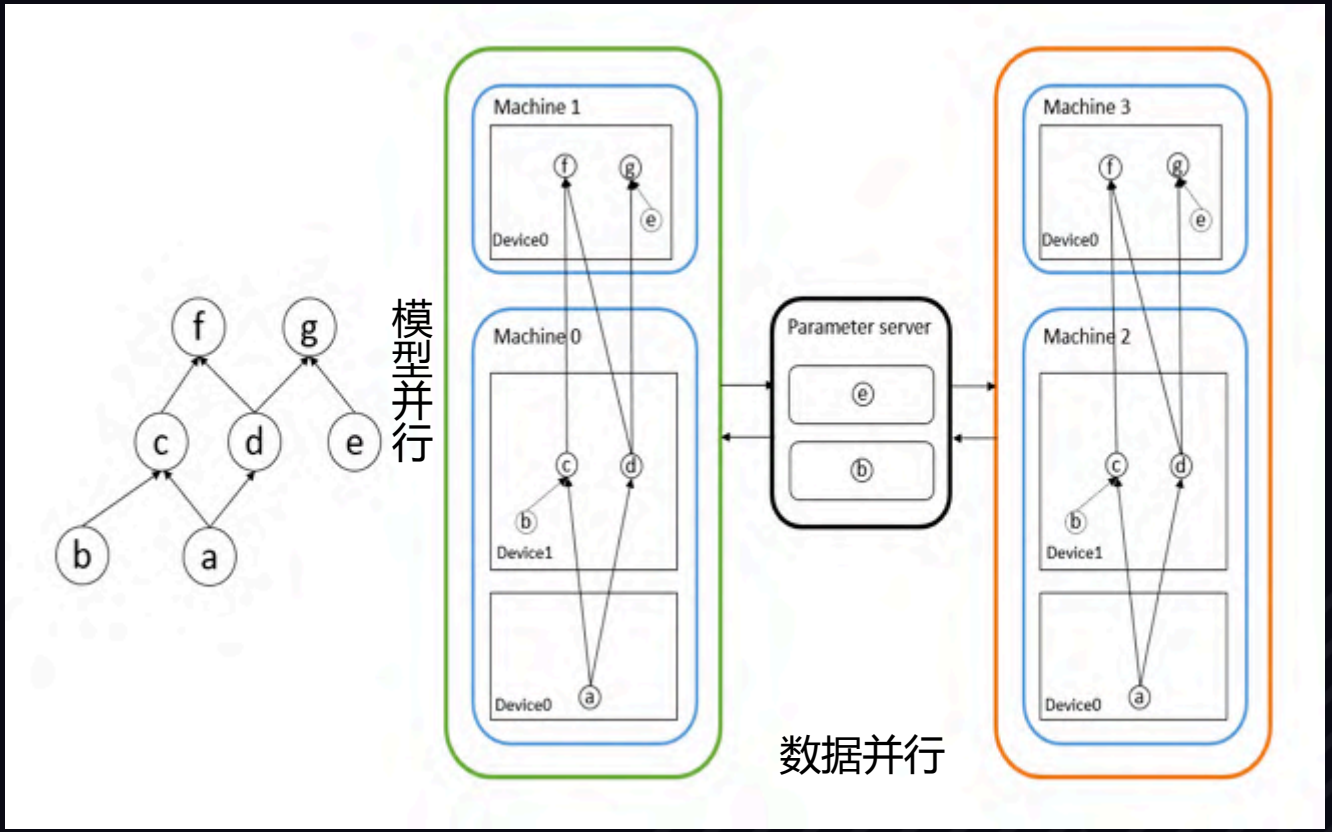

模型训练
手动训练
新建
选择数据源
数据预处理
特征选择
训练
测试
评估
自动训练
自动
获取数据
数据预处理
特征选择
训练
测试
评估
推送模型上线
模型运行
模型服务接口暴露
同步
异步
模型统一运算服务
数据预处理
变量计算
模型运算
运行
结果统计
模型加载
模型
生命周期
模型管理
模型集群管理
训练任务
管理
模型运行配置
模型生命周期
模型算法管理
算法
SVM
Bayes
回归
决策树
RF
GBDT
CNN
...
模型评估
变量评估
运行统计报表
模型评估
衰退评估
模型存储
训练任务存储
运行变量快照存储
运行分值存储
运行统计信息存储
MySQL
云梯
HBASE
OSS

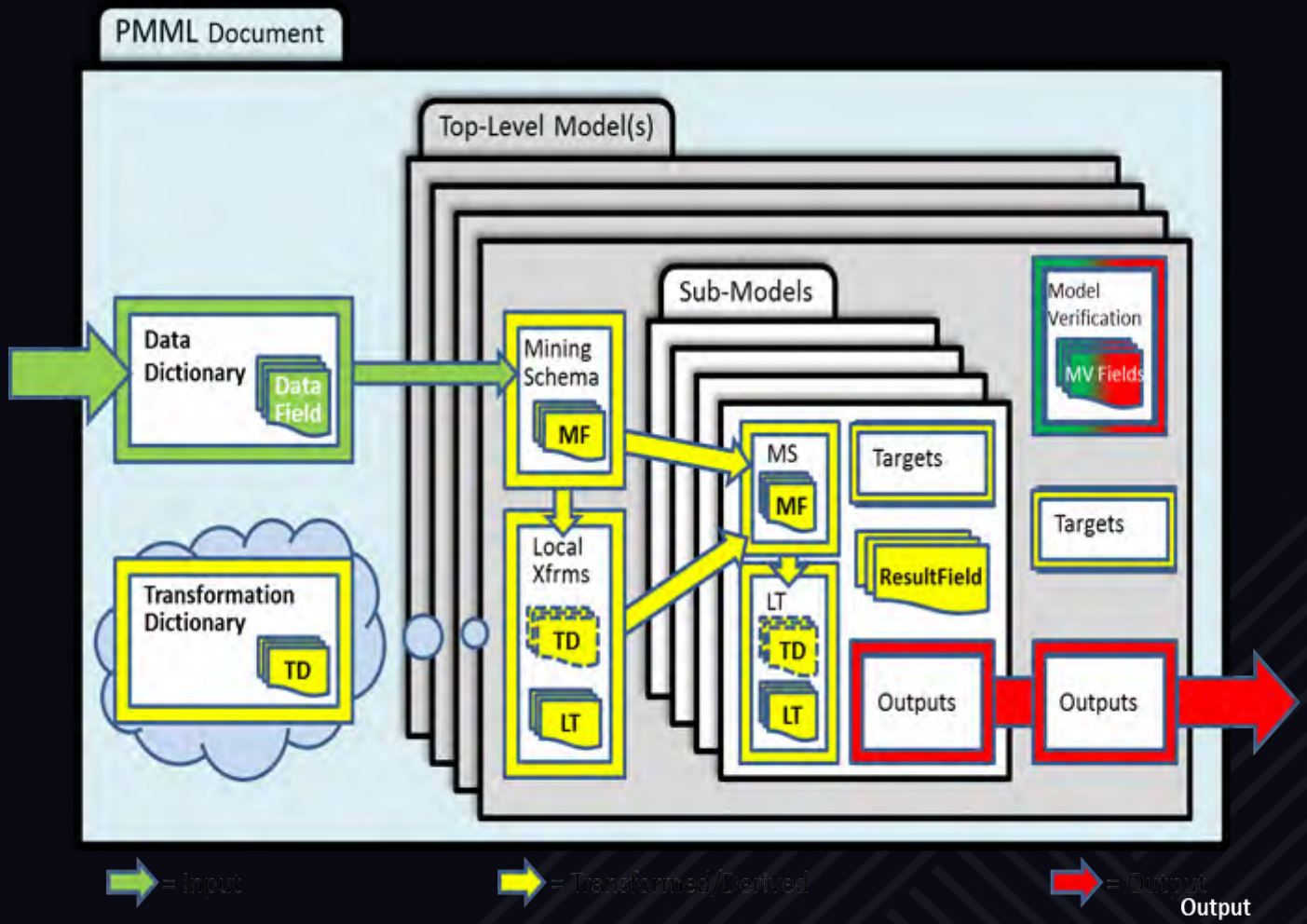
TABLE OF
CONTENTS 大纲

架构经验：分布式

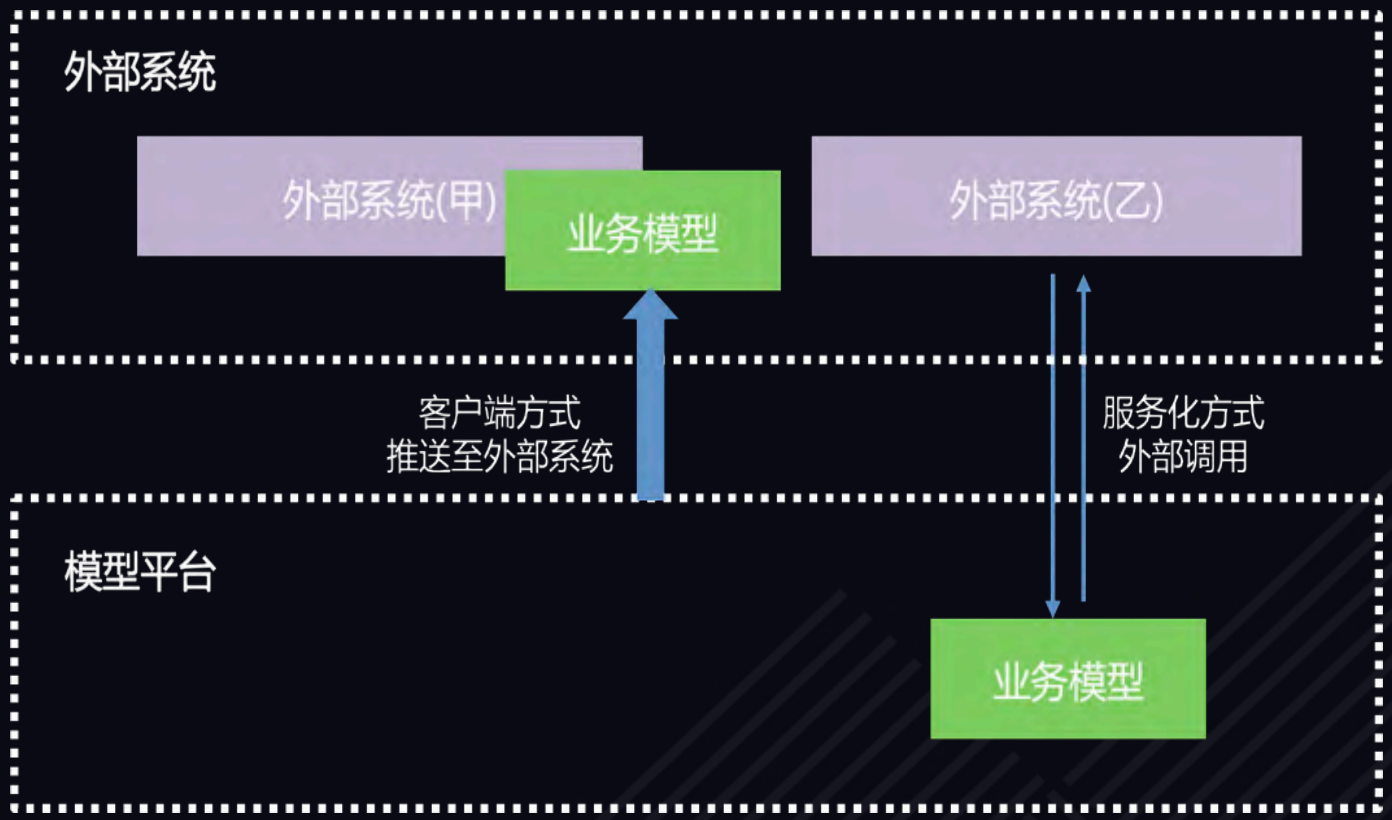

f
g
c
d
e
b
a
模型并行
Machine 1
Device0
Machine 0
Device1
Device0
Parameter server
e
b
Machine 3
Device0
Machine 2
Device1
Device0
数据并行

分布式

架构经验：PMML


PMML Document
Top-Level Model(s)
Sub-Models
Data Dictionary
Data Field
Transformation Dictionary
TD
Mining Schema
MF
Local Xfrms
TD
LT
MS
MF
LT
TD
LT
Targets
ResultField
Outputs
Model Verification
MV Fields
Targets
Outputs
= Input
= Transformed/Derived
= Output
Output

架构经验：性能



复杂场景：客户端+服务端，分层性能优化

架构经验：特征预热

新特征的实验与预热

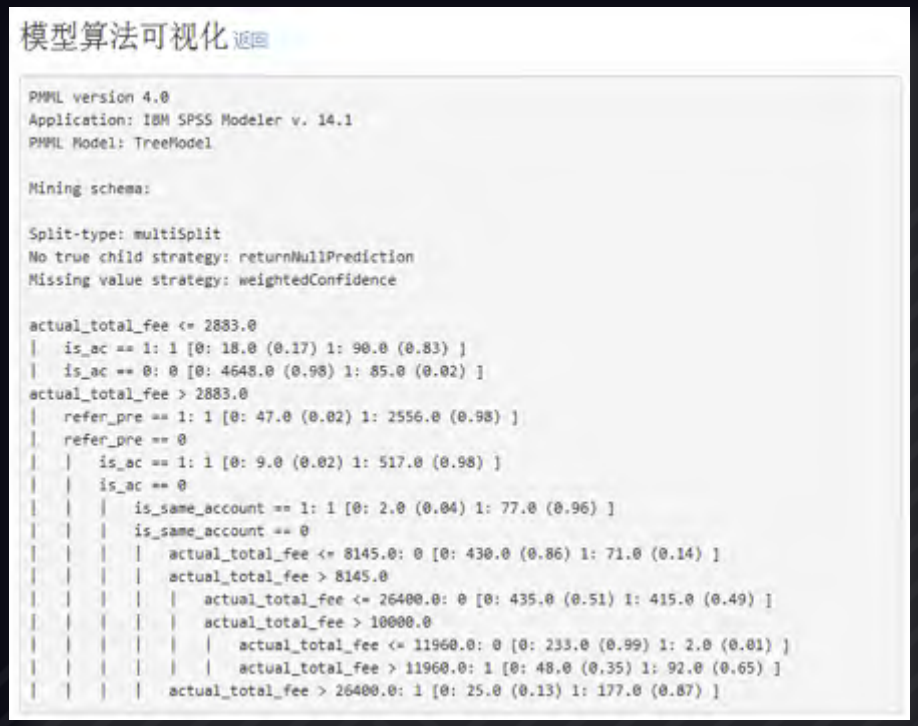
特征定义上线：
定义 与 加工过程（UDF/ Online函数）的同步

数据上线：
融合计算：离线数据与实时数据融合

模型外化与可解释性

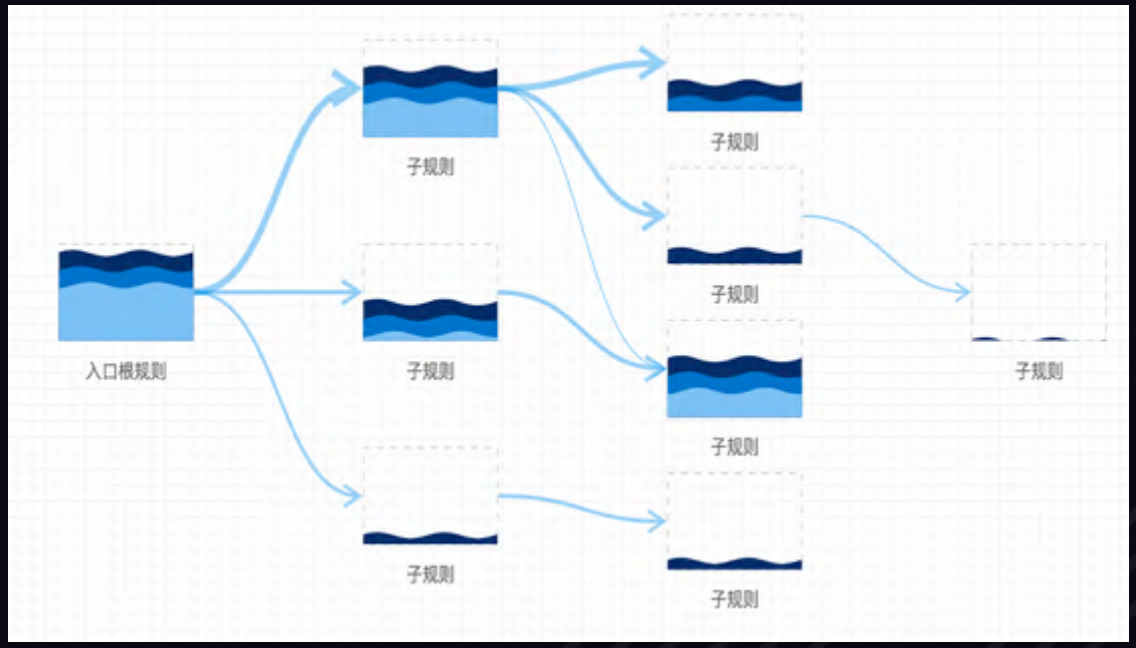
变量可视化

算法可视化



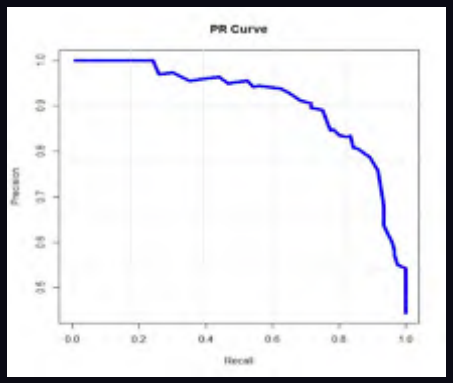
深度学习的可解释性?

模型外化与可解释性



规则化展示，提高可解释性

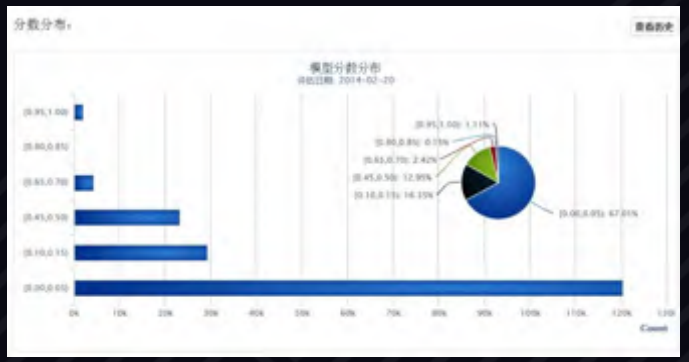
模型衰退与评估



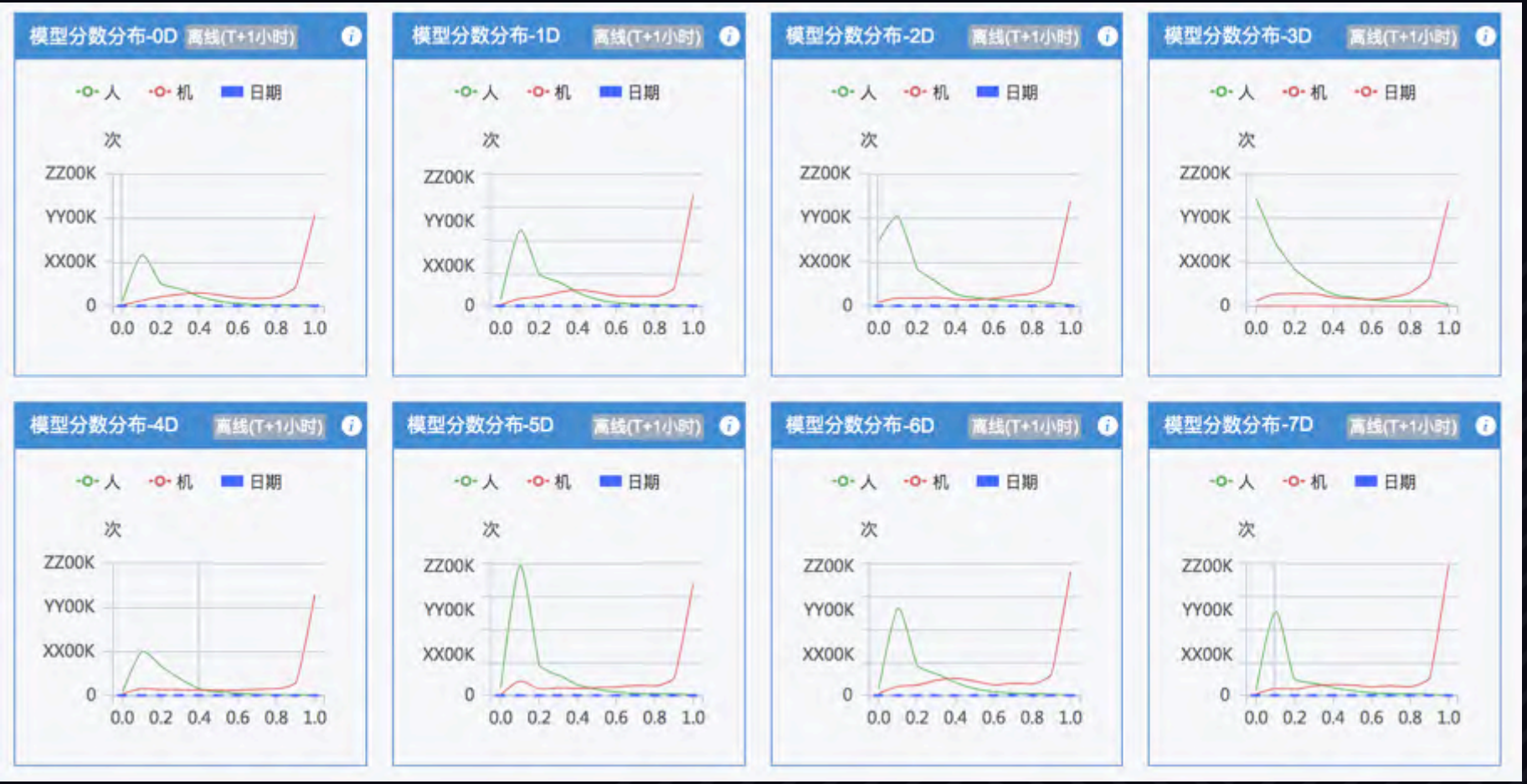
变量评估

变量	总数	空数	详细
[redacted]	179464	0	0 99.68% 1 0.32%
[redacted]	179464	0	0 97.28% 1 2.72%
[redacted]	179464	0	0 99.67% 1 0.33%
[redacted]	179464	0	Avg 4692 Min 35 Max 61600000 方差 21330742301

Actual	Predicted		
		Positive	Negative
	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)



模型衰退与评估


模型分数分布-0D 离线(T+1小时)
模型分数分布-1D 离线(T+1小时)
模型分数分布-2D 离线(T+1小时)
模型分数分布-3D 离线(T+1小时)
模型分数分布-4D 离线(T+1小时)
模型分数分布-5D 离线(T+1小时)
模型分数分布-6D 离线(T+1小时)
模型分数分布-7D 离线(T+1小时)
人
机
日期
次
ZZ00K
YY00K
XX00K
0
0.0 0.2 0.4 0.6 0.8 1.0

组合模型

- ✓ 从不同维度学习（约束各特征影响范围）
- ✓ 发挥不同学习器的特长
- ✓ 子模型可单独维护与评估
- ✓ 便于新特征加入

架构

没有最完美的架构，只有最合适的架构

架构是演进出来的，而不是设计出来的

THANKS!

让创新技术推动社会进步

HELP TO BUILD A BETTER SOCIETY WITH
INNOVATIVE TECHNOLOGIES

Geekbang.
极客邦科技

InfoQ
专注中高端技术人员的技术媒体

EGO EXTRA GEEKS' ORGANIZATION NETWORKS
高端技术人员学习型社交平台

StuQ
斯达克学院
实践驱动的 IT 教育平台

地址：北京市朝阳区洛娃大厦C座8层1801室　　网址：www.geekbang.org