阿里云机器学习技术与应用

主讲人：阿里云高级专家 刘吉哲

由BI转向机器学习

- 数据化运营需求公司已达200万家（来自国家统计局）
- 约20万家企业意识到机器学习的价值
- 2016-2018约 60% 大中型公司将机器学习用于商业分析

技术门槛高

数据量大

硬件成本高

机器学习开发面临的挑战

阿里云机器学习 PAI

高性能云端计算

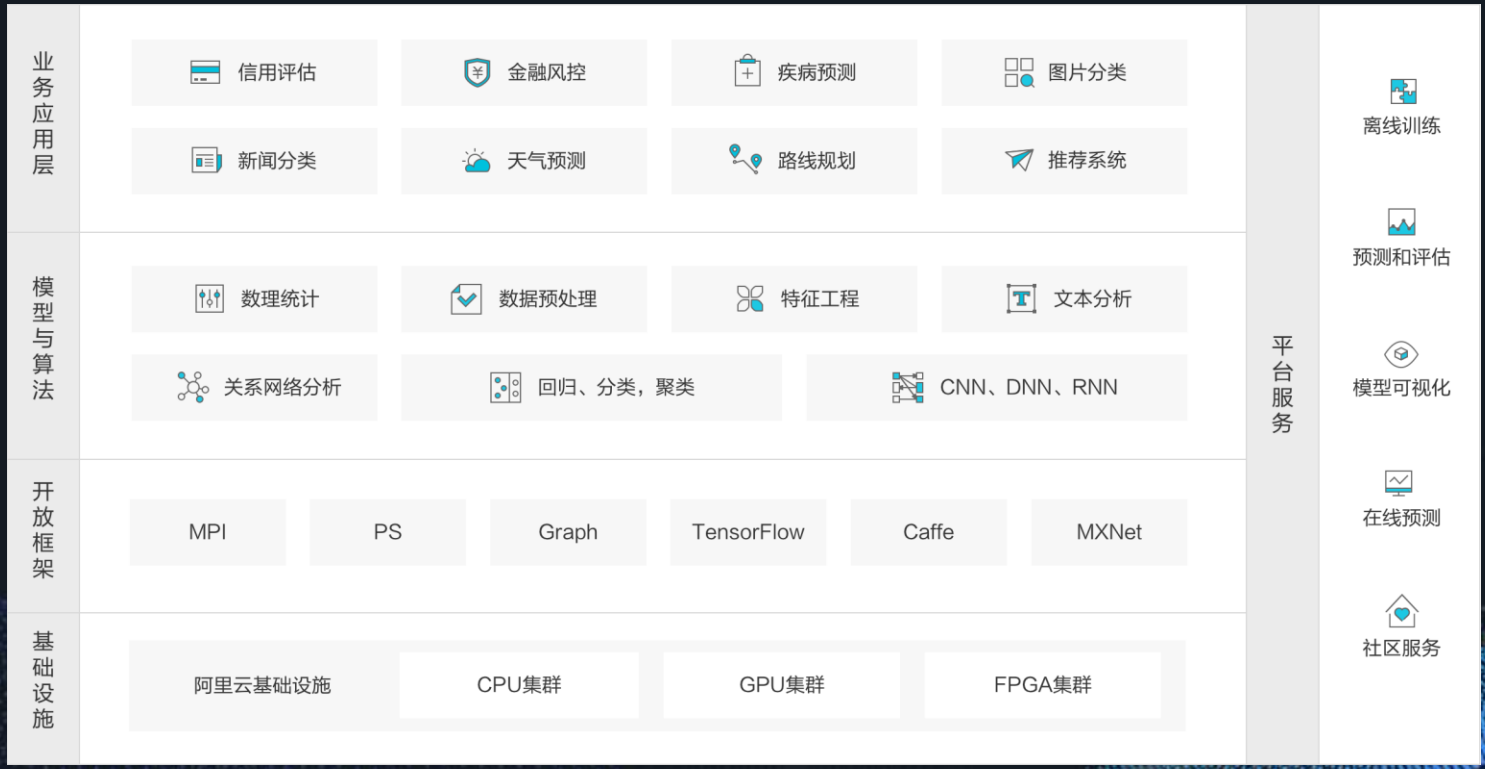
降低存储和计算成本

工具、算法库

降低技术门槛

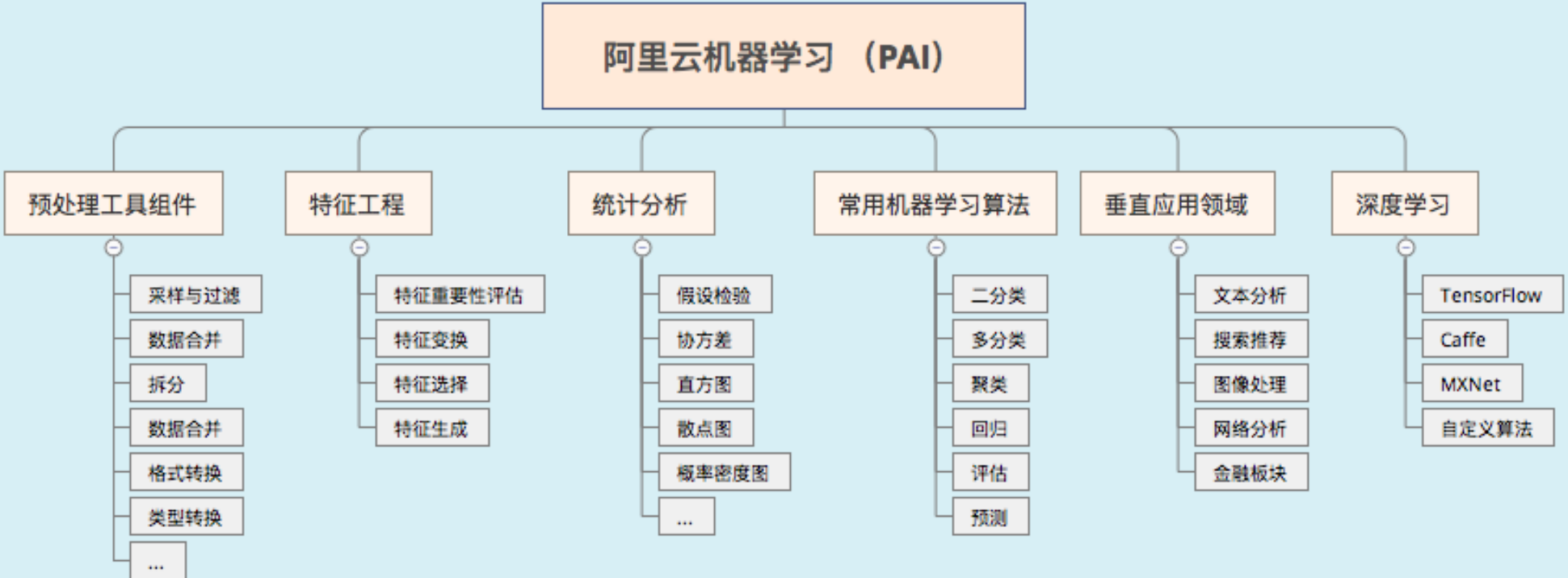
目录
Contents

阿里云机器学习 PAI 2.0




机器学习平台新增tensorflow、mxnet、caffe学习框架!
最新功能
深度学习DNN算法
云栖社区功能
新手引导
PAI产品概述
新手入门
常见问题
F&Q
客服在线
云栖社区
【玩转数据系列二】机器学习应用没那么难，这次教你玩心脏病预测
首页
实验
数据源
组件
模型
设置
更多
基础
新建空白实验
开发流程
数据
数据预处理
特征工程
机器学习模型训练
模型评估
学习
离线/在线服务
【推荐算法】商品推荐
推荐
农业贷款预测的回归算法...
shequde...
工作空间
简体中文

算法种类丰富



算法脱胎于内部业务

01 MPI 算法和SDK
定向广告，微贷风控

02 PS 算法和SDK
淘宝主搜推荐系统

03 深度学习算法
芝麻信用、蚂蚁证件审核

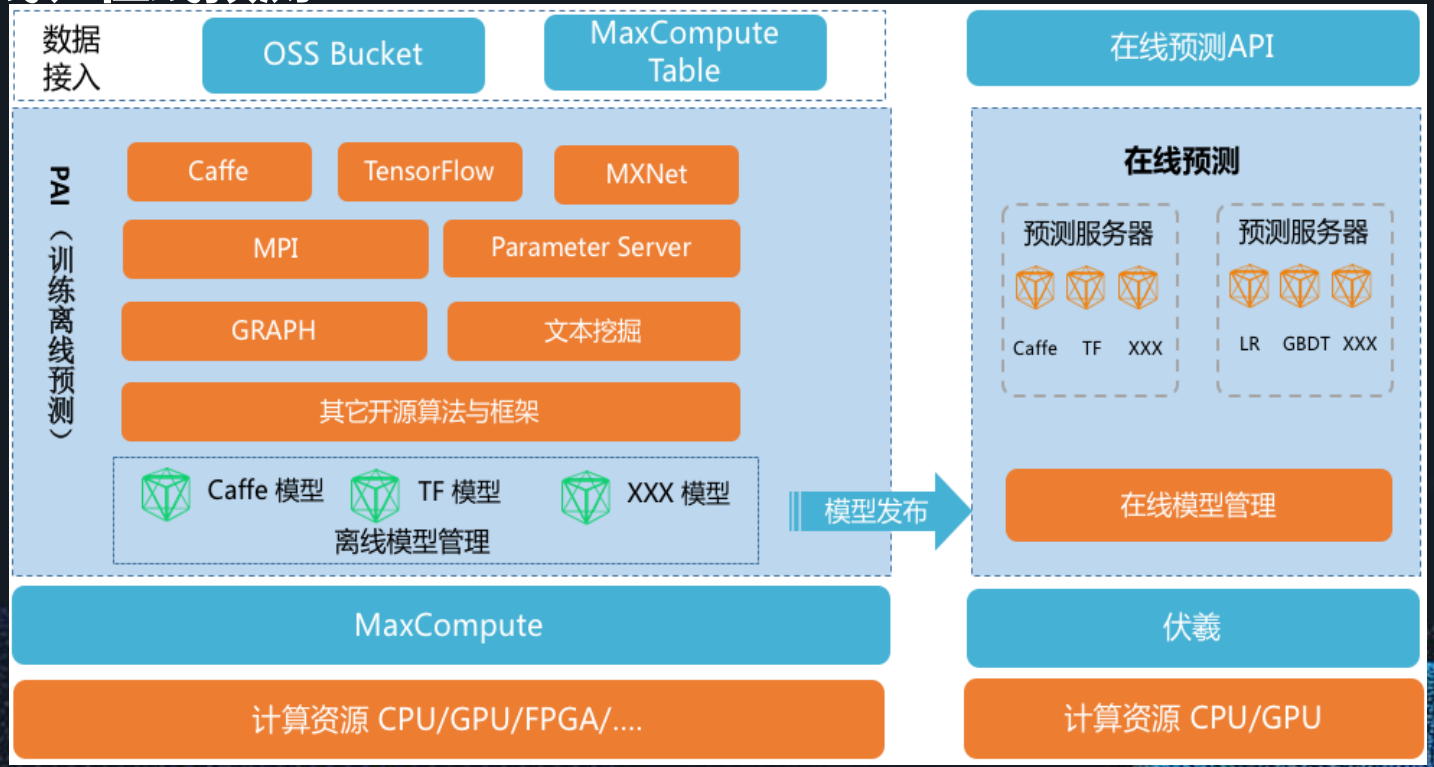
深度学习框架

1 **TensorFlow**
支持版本 1.0

2 **Caffe**
支持版本 RC3

3 **MXNet**
支持版本 0.9.5

模型离线、在线预测

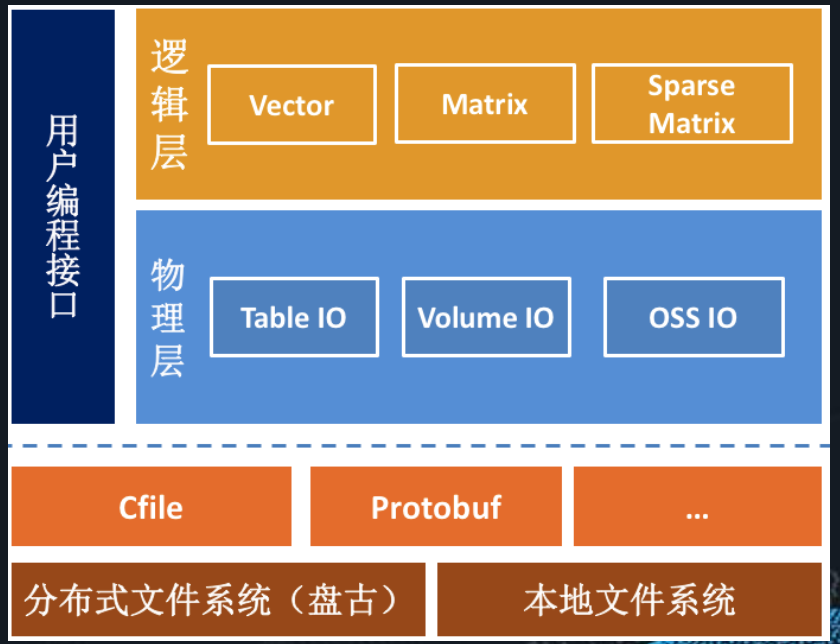


目录

Contents

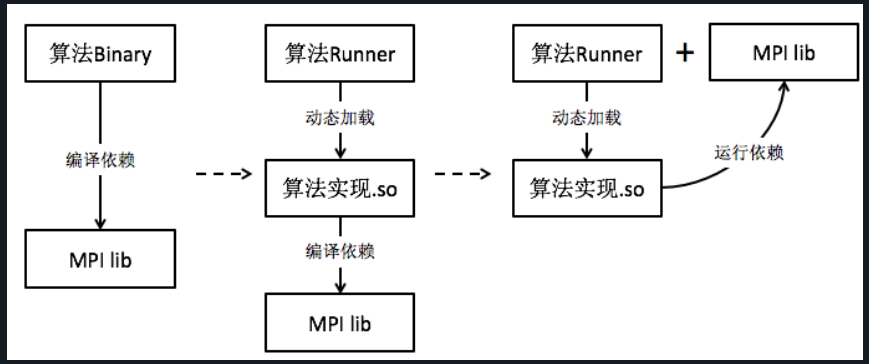
支持结构化、非结构化数据

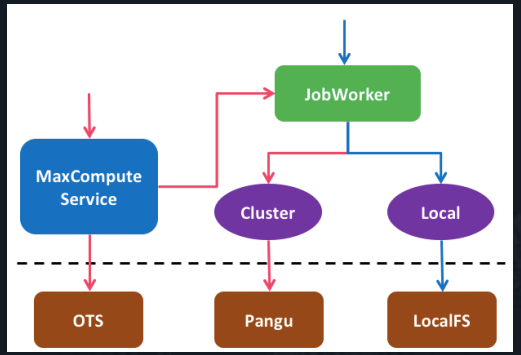
- 结构化
  - MaxCompute 表
    - 数据是无序的
    - 特征表达有限
  - KV表
  - Map表
    - SQL select支持
- 非结构化
  - 图像、语音、视频
  - 支持向结构化数据转换



MPI 算法开发

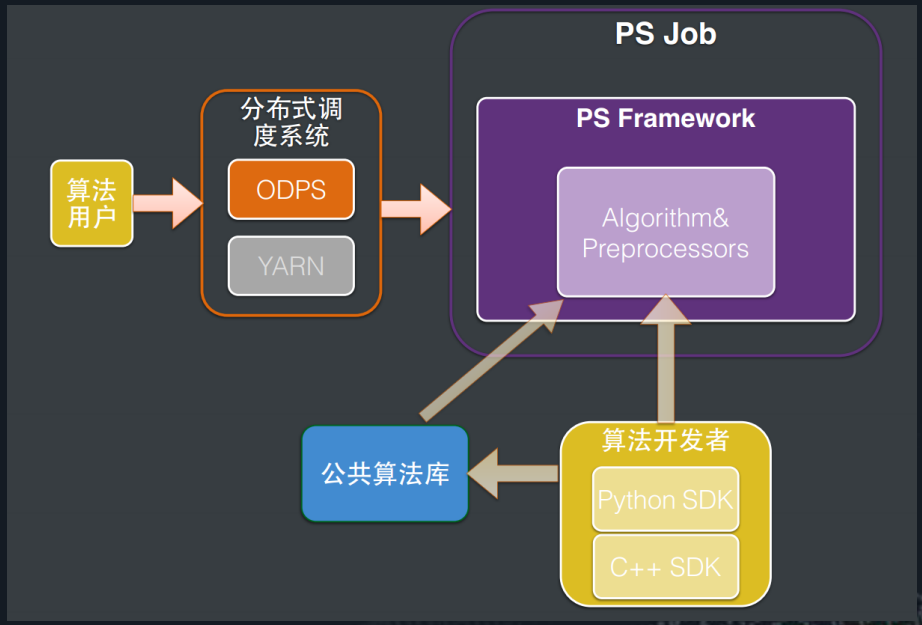
- 开发
  - 平台与算法解耦
  - MPI 计算框架与算法分离
- 调试
  - 本地开发、调试
  - 线下、线上无缝集成





参数服务器/PS

- 参考实现
  - Google DistBelif
  - Limu Parameter Server 架构
- 广告推荐
  - 100T，260亿特征
  - 8个小时收敛
- PS-SDK接口
  - Python/C++混合编程
  - 提供本地调试

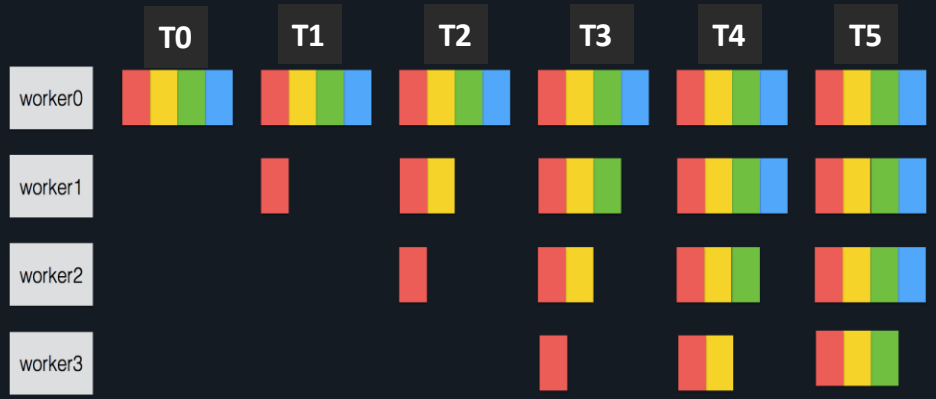


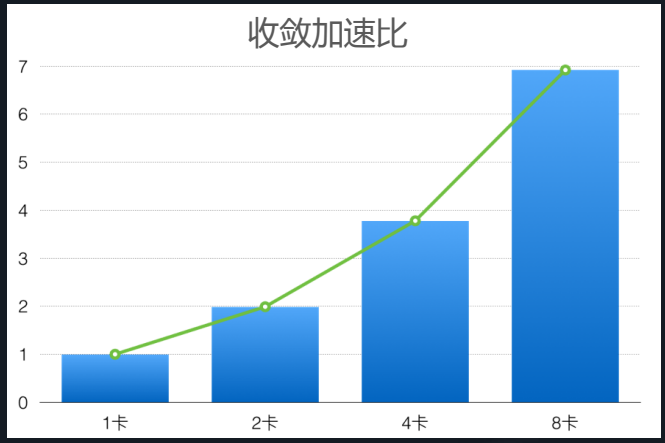
文本分析

文本分析算法（API）

- 词频统计
- TF-IDF
- PLDA
- Word2Vector
- Doc2Vector
- 关键字抽取
- 文章相似度
- 文本摘要
- 条件随机场预测

Caffe 多机多卡





AlexNet On ImageNet Data

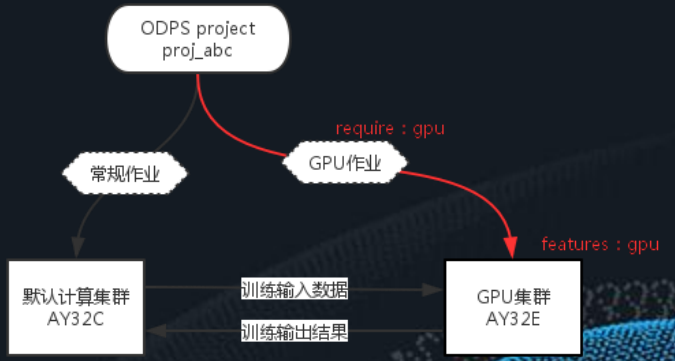
- 数据分片，1-N 变为 P2P
- AlexNet 8卡上得到将近7倍性能提升

异构(CPU/GPU)集群调度

- Job DAG 中允许不同Action类型节点

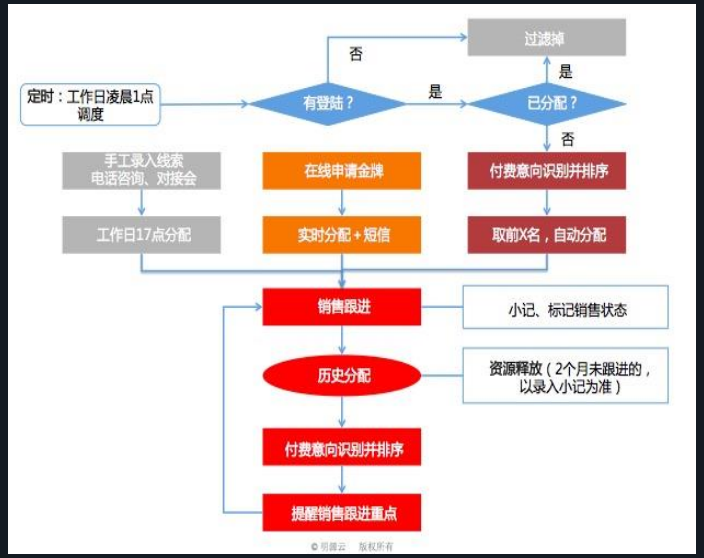
- 扩展Job XML，基于 tag 确定运行集群

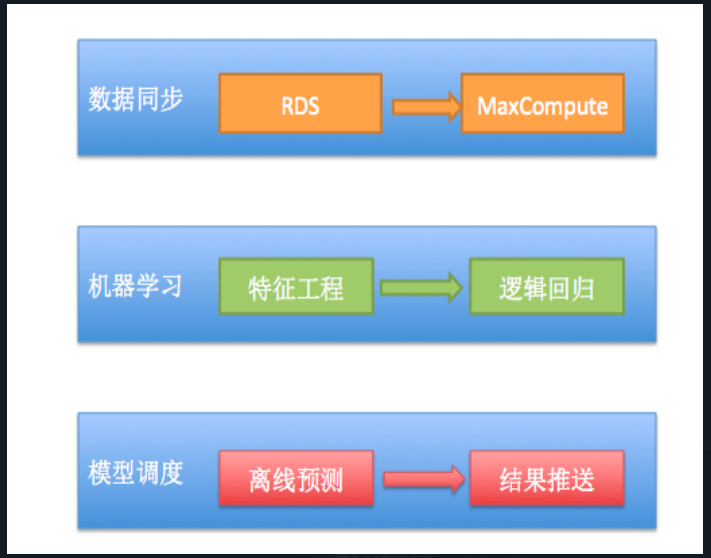
```
<resource>
  <instance_count>3</instance_count>
  <cpu>100</cpu>
  <mem>4000</mem>
  <virtual_resources>
    <r name="GPU" value="100" />
  </virtual_resources>
</resource>
```



目录
Contents

明源地产



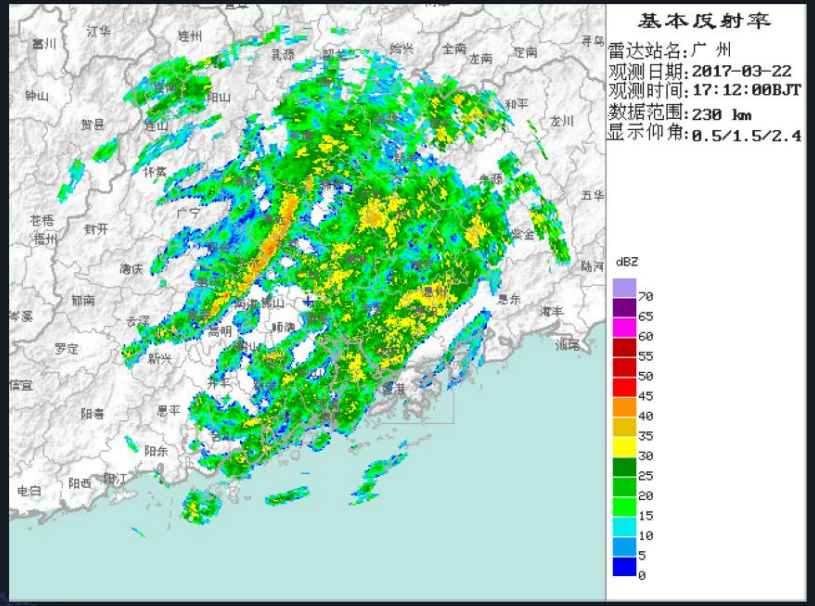


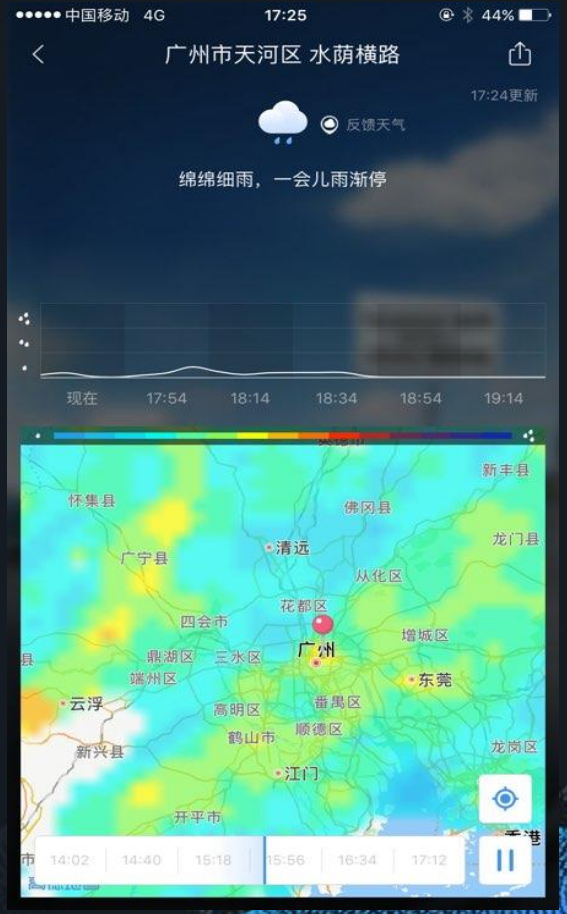
- 在业务流程中融入机器学习
- 付费客户转化比由10.7% 提升到 19.6%

广告无打扰植入

- 自动植入式广告，不影响观看体验，不停顿、无打扰
- 可根据视频场景和观众喜好植入定制化广告内容

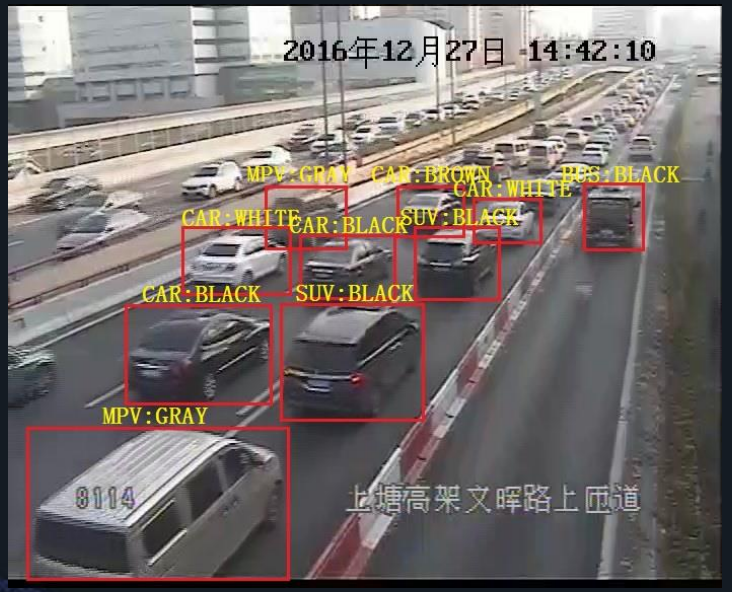
墨迹天气





- 根据雷达回波图样本训练生成云图模型
- 通过在线预测实现区域短时天气预报

杭州城市大脑



- 根据车型和品牌，计算每辆车所在位置和物理距离，计算通行速度
- 根据视频数据，进行异常检测，实时预警，提升事件处理效率

天池算法大赛(Power By PAI)

2016

- 菜鸟分仓规划
- 阿里音乐流行趋势预测
- 最后一公里极速配送
- 阿里云安全算法挑战赛
- 机场客流量时空分布预测

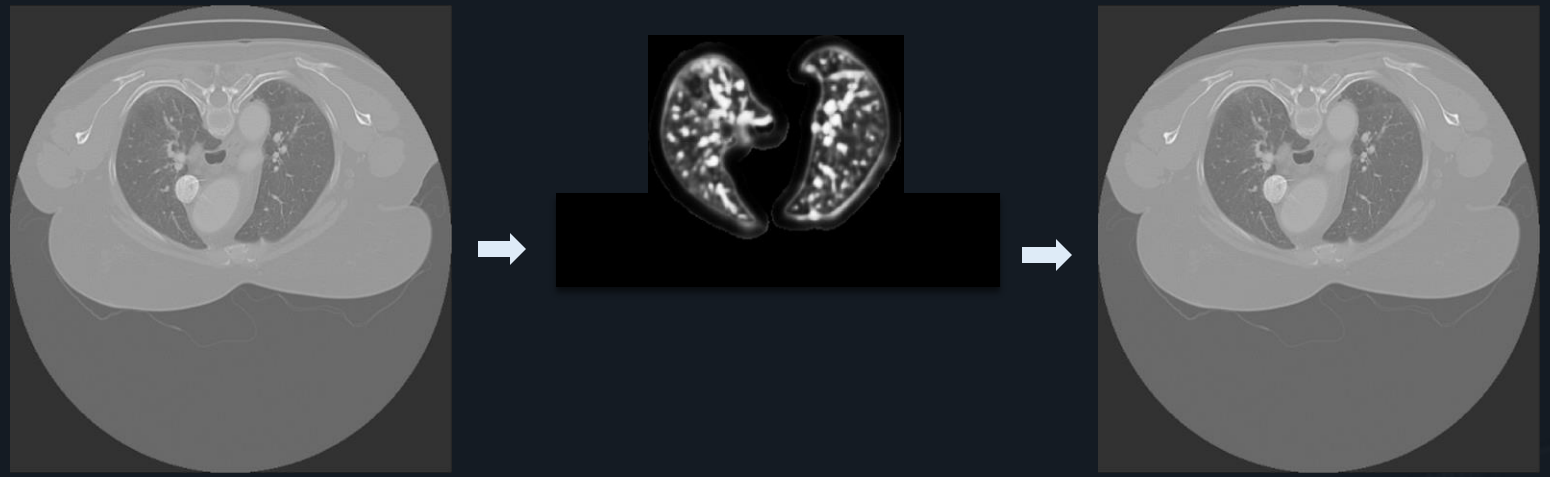
2015

- 阿里移动推荐算法
- 资金流入流出预测
- 新浪微博互动预测
- 淘宝穿衣搭配算法
- 公交线路客流预测

2017

- IJCAI-17 口碑商家客流量预测
- KDD CUP 2017
- 阿里聚安全算法挑战赛
- 更多赛事……

医学影像分析



- 阿里云与 Intel 合作，7月份在 PAI 产品上开展天池医学影像分析算法大赛

总结

- 机器学习与深度学习相结合的平台
- 强大的计算资源
- 一站式人工智能机器学习平台

更多学习资料与支持请扫二维码进入

阿里云机器学习 PAI 官网

数加•阿里云机器学习钉钉群


2017云栖大会·成都峰会
THE COMPUTING CONFERENCE
阿里云
云栖社区
yq.aliyun.com
飞天·智能
APSARA INTELLIGENCE
2017云栖大会·成都峰会
5月23日 成都世纪城天堂洲际大酒店