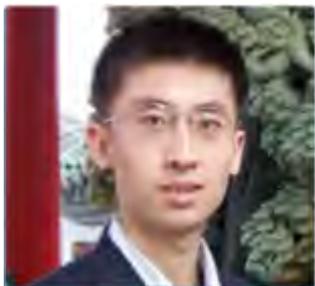




Aura - 以机器学习为核心的 数据驱动型应用开发平台

AsiaInfo Data Science Team

亚信数据科学团队



常剑, Ph.D. from University Of Pennsylvania

亚信数据硅谷研发中心 高级数据科学家

数据驱动产品的核心技术研发

BIOGRAPHY

宾夕法尼亚大学 (University Of Pennsylvania) 计算机博士 清华大学本科

专注于大数据和机器学习技术的研发工作，博士期间进行大规模分布式系统的安全与信任问题的研究。研究成果在国内外的顶级期刊和会议如：ACM Computing Surveys, ACSAC, CEAS, EuroSec, FGCS, HiCoNS, HSCC, IEEE Systems Journal, MASHUPS, PST, SSS, TRUST, and WiVeC 等发表众多的文章。

领导技术团队与中信银行成立业内第一个金融科技创新实验室，进行支付欺诈防控，智能投顾等新产品的研发。同时也是诸多国际著名期刊和会议如 ATVA, CORCS, EMSOFT, ETFA, FM, ICCAD, ICCPS, ICDE, ICECCS, IET-IFS, INS, ISORC, JCSE, RTSS, SRDS, SJS, SSS, SWJ, TPDS, WCL, WH 的审稿人。

About Us



徐争, Ph.D. from Technology University Of Delft

亚信数据数据科学家

机器学习与深度学习算法研发

BIOGRAPHY

荷兰代尔夫特大学 (Technology University Of Delft) 天体动力学博士
欧洲空间局 (European Space Agency) 数据科学家

多年从事机器学习算法的研究以及在数据分析、建模方面的应用。涉及的行业横跨天体动力、卫星遥感、气候变化、海洋以及冰冻圈/水圈等多个科研领域。相关研究成果发表于国外顶级期刊与会议如: Geophysical Journal International, The Cryosphere, Journal of Geophysics Research , Journal of Hydroinformatics, AGU, EGU等。

加入亚信后专注于人工智能技术在公共安全、物联网以及大数据领域的产品研发。相关产品利用深度学习解决了对视频数据进行对象分类、识别、检测以及跟踪的问题，例如：人流/车流管控等。另外在与深圳司法局的联合科研项目中，利用视频数据以及机器学习技术提升了监狱安防系统日常监管的智能化，得到了司法局领导的认可。作为亚信数据数据科学团队的一员，目前正在参与研发新一代面向运营商的机器学习与深度学习通用平台 (Aura)。

亚信数据—中国电信行业第一大数据域软件提供商



1993

亚信在美国成立



2000

亚信在美国NASDAQ上市
股票代码：ASIA



2010

收购南京联创
成为中国第一、全球第二的电信软件供应商



2014

亚信私有化

亚信数据从集团分拆
数据分析产品跨行业突破



2015

亚信数据正式成立

亚信数据

- 中国通信行业第一大数据域软件提供商
- 2014年8月业务从亚信集团分拆，员工2000+
- 中信资本占约1/3的股权份额



10亿 承载电信用户数



8000T 数据系统每日分析量



10亿 2015年软件订单

OVERVIEW

目录

1 Retrospective and Looking Forward

大数据与人工智能技术的回顾与展望

2 AURA - AsiaInfo Machine Learning Platform

Aura 数据驱动型应用开发平台简介

3 Data-Driven Application Use Cases

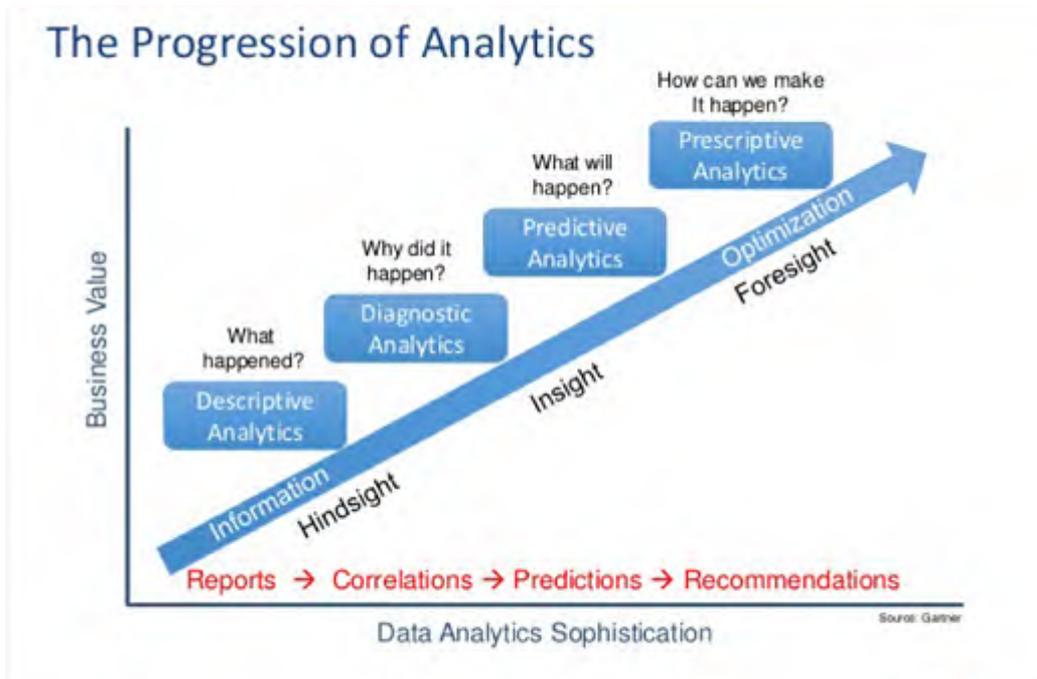
数据驱动型应用示例

4 Discussion

问题与讨论

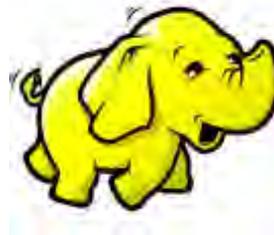
Applications of Big Data

- Examine the past
 - ▣ Descriptive Analytics
 - ▣ Diagnostic Analytics
- Now and future
 - ▣ Predictive Analytics
 - ▣ Prescriptive Analytics



Examine the Past

- Ad-hoc
 - ▣ Data warehouse – OLAP
- Pre-computed
 - ▣ Batch (1st gen)
 - Hadoop for batch processing and reporting
 - ▣ Lambda (2nd gen) – dual pipeline
 - Hadoop for batch processing and reporting
 - Storm / Spark Streaming / Samza for real-time processing
 - ▣ Kappa (3rd gen) – single pipeline
 - Real-time computation and query / reporting
 - Fault-tolerant distributed real-time architecture
 - Re-playable, real-time event streams



Now and Future

- Predictive Analytics
 - ▣ What will happen?
 - ▣ Predict the future outcome based on historical events
- Prescriptive Analytics
 - ▣ Actionable - actions to be performed due to the predictions and showing the implications of each decision option
- Machine learning and deep learning
 - ▣ Customer acquisition (telecom, banking, e-commerce, ...)
 - ▣ Fraud detection (telecom, banking, healthcare, advertising, ...)
 - ▣ Customer churn prevention (telecom, banking, ...)
 - ▣ Object recognition, facial recognition, ...

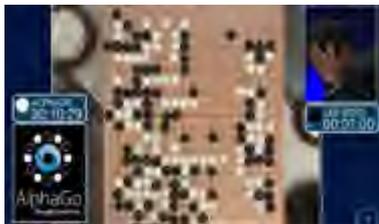


Defining Artificial Intelligence

定义人工智能



2016年3月15日，在人机大战最终局结束后，李世石独自一人掩面坐在棋盘前。这场比赛主动挑战执黑的李世石没有能再次展现出“神之一手”，最终180手投子不敌AlphaGo。而人类VS人工智能 (AI, Artificial Intelligence) 也最终以1比4落下大幕……



✓ AlphaGo计算集群

- 1920 个CPU
- 280 个GPU

✓ 耗时数月完成模型训练

What is AI ?

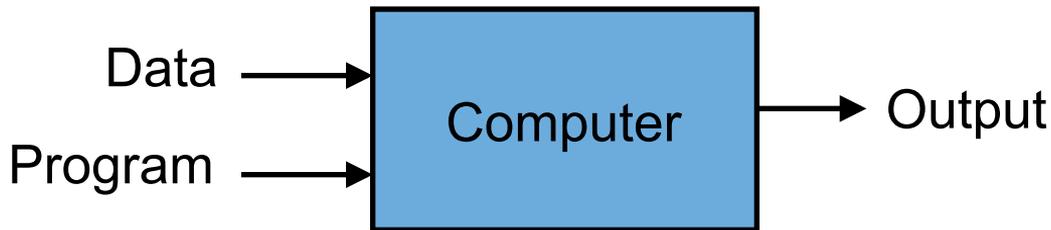
- Artificial intelligence (AI) is intelligence exhibited by machines.
- The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.
- In computer science, the field of AI research defines itself as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal.

Machine Learning is an important area of AI 机器学习是人工智能的重要领域

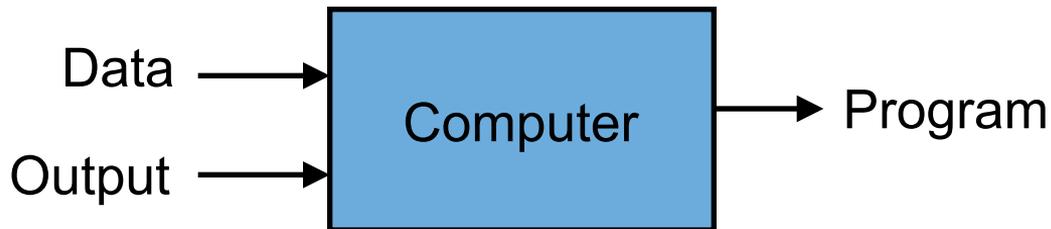
- Machine Learning is about:

- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck, let the data do the work instead!

Traditional Programming



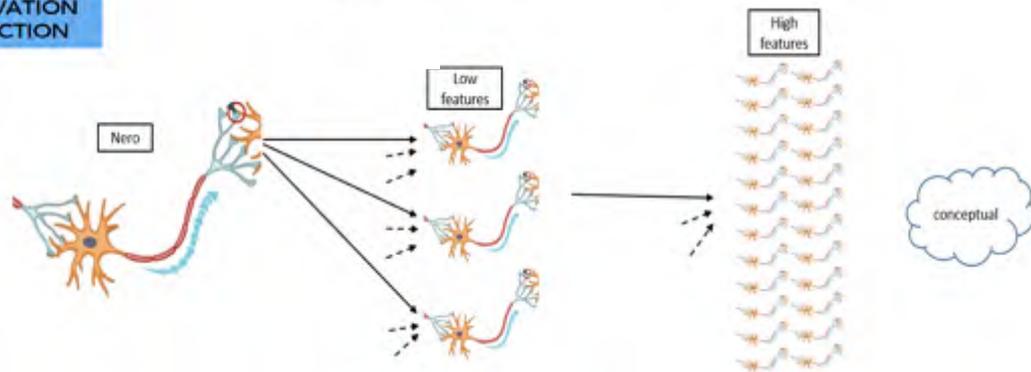
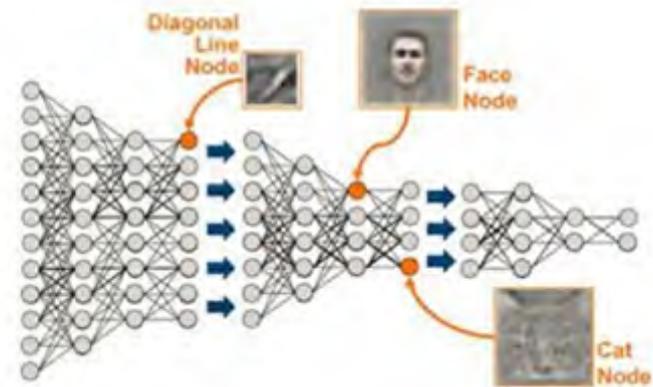
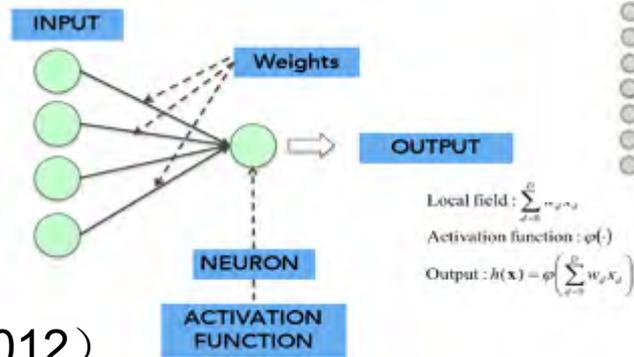
Machine Learning



History Progress in Deep Learning 深度学习的历史进程

- ✓ Artificial neural network is the building block
- ✓ Mimic human brain
- ✓ High depth and width
- ✓ Heterogeneous computing (CPU + GPU)

- Google “Brain” (2012)
- IBM “Watson” (2015)
- 百度大脑 (2015)
- NVIDIA DGX-1 (2016)



AisaInfo is a Pioneer in AI



行业
及业务范畴

亚信数据

亚信软件

亚信安全

亚信国际

运营商
Telco

医疗
Healthcare

银行
Banking

媒体
Media

公安
Public Safety

...

智能营销

流失预警

...

医疗控费

医疗经验传播

...

个人客户
智能投顾

支付欺诈检测

...

内容智能推荐

收费会员画像

...

公证业务预测

公证人群画像

...

...

...

...

AI算法
及模型知识库

智能推荐算法
及模型知识库

流失预警算法
及模型知识库

用户画像算法
及模型知识库

骗保预测算法
及模型知识库

热点业务算法
及模型知识库

...

团队核心能力

业务咨询能力

自然语言
处理能力

语音及图像识
别能力

数据挖掘能力

新技术研究及
业务转化能力

...

OVERVIEW

目录

1 Retrospective and Looking Forward

大数据与人工智能技术的回顾与展望

2 AURA - AsialInfo Machine Learning Platform

Aura 数据驱动型应用开发平台简介

3 Data-Driven Application Use Cases

数据驱动型应用示例

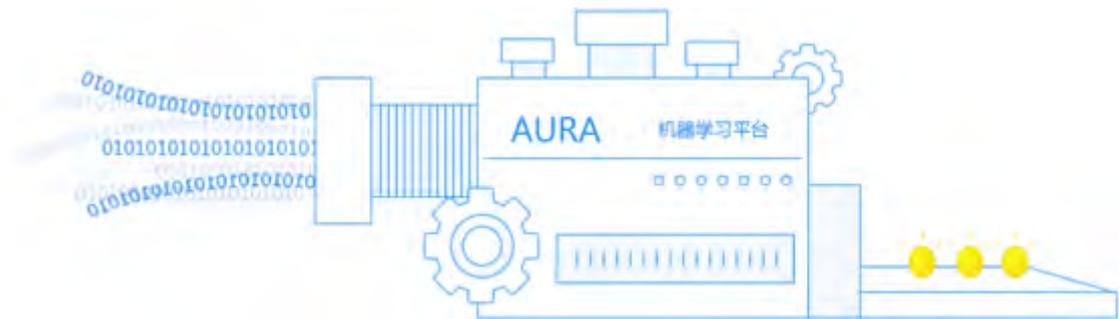
4 Discussion

问题与讨论

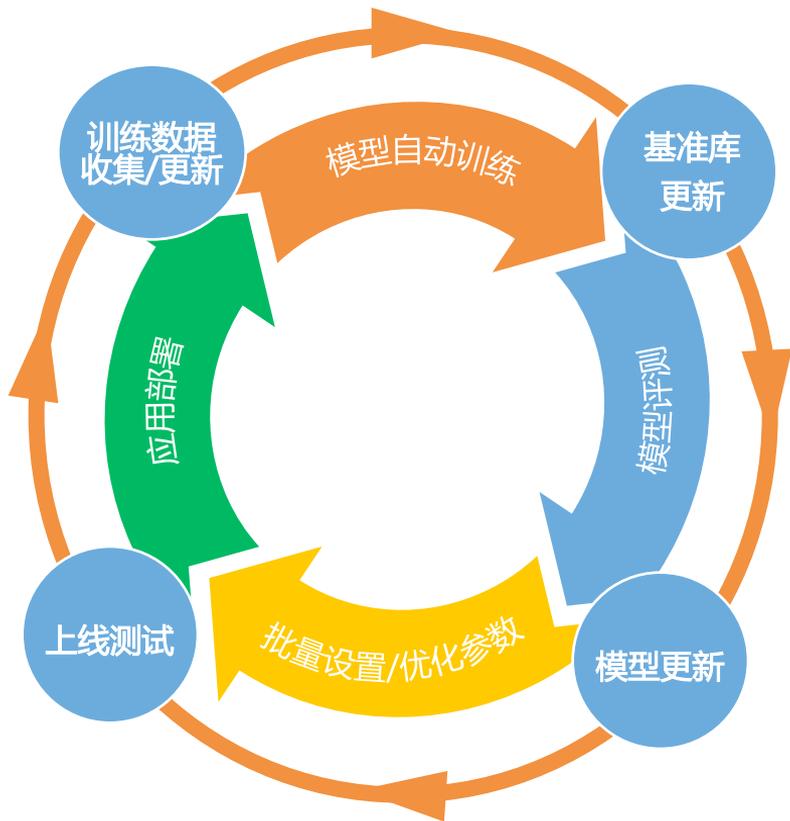
Aura Positioning

Aura是基于机器学习和深度学习技术的**数据驱动型应用**的**高效开发平台**。用户能够:

- 应用行业数据应用已有的最佳实践
- 采用图形化或编辑器两种模式开发
- 沉淀企业私有的数据挖掘的探索成果

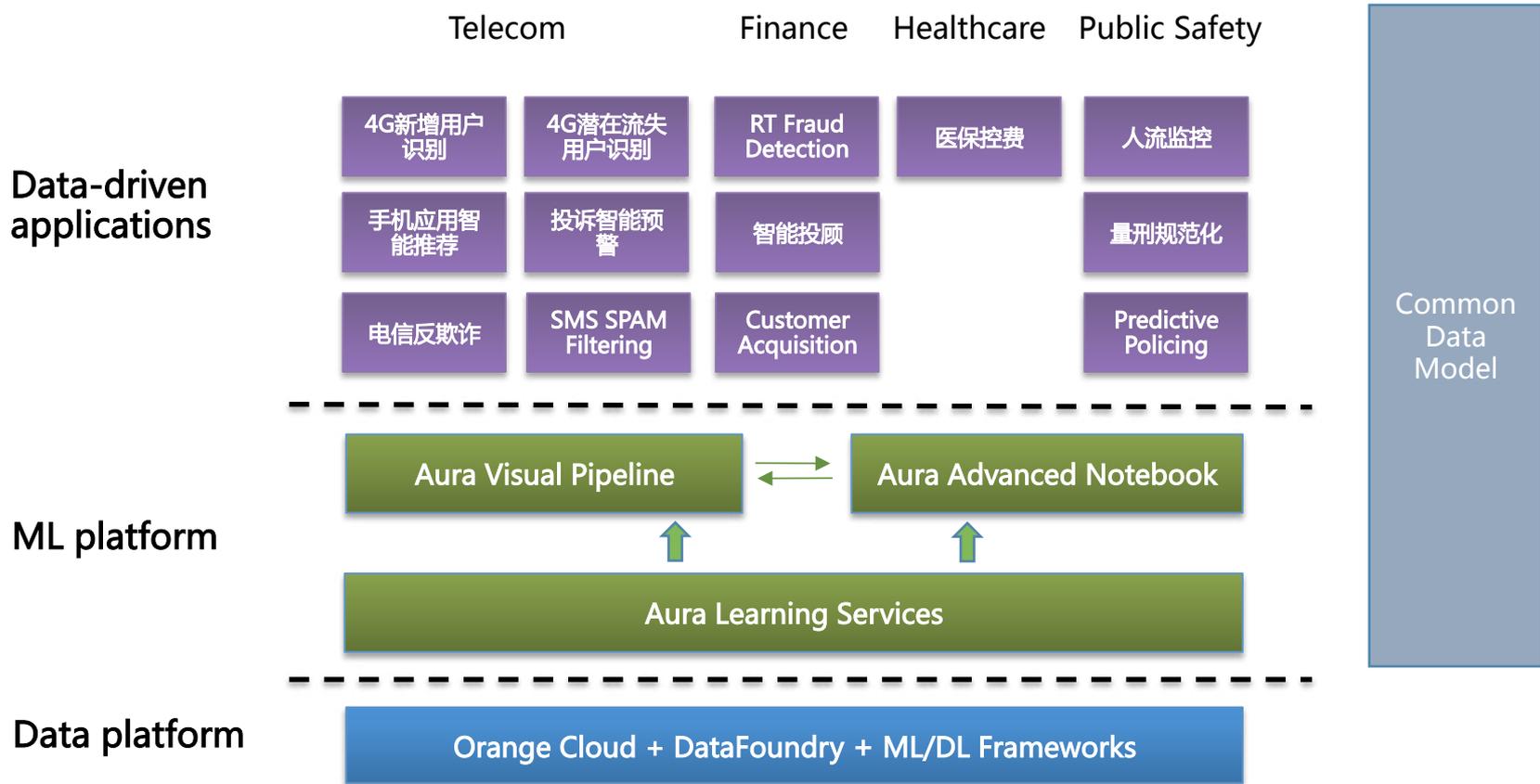


Aura - System Reinforcement Loop

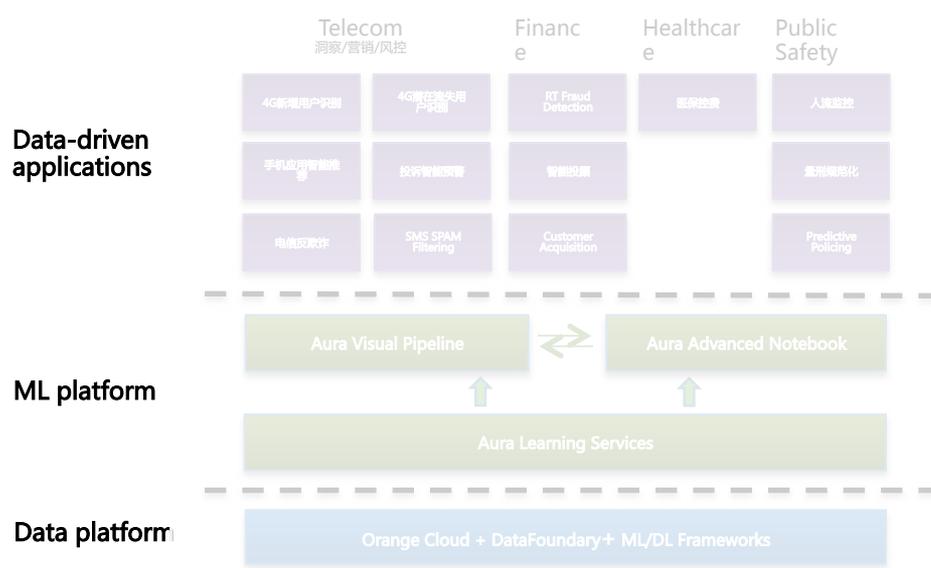


- ✓ 行业数据标准积累，技术与业务知识的充分结合
- ✓ 预测结果与业务数据积累在同一平台形成闭环
- ✓ 模型自优化，自动评价模型表征，学习新样板数据中的特征变化，更新模型，提升模型准确度；
- ✓ 根据实际效果自优化，根据每一轮实际的效果评估数据，自动调整模型，不断提升准确度。

Aura Architecture



Aura - Common Data Model



AsiaInfo
Common
Data
Model
(CDM)

Common Data Model

- 行业数据模型的定义
 - 字段、类型、语义、关系
- 定义不同数据模型之间的关联关系
- 提供上层数据调用的接口和辅助操作
- 行业数据标准和数据质量的校验
- 不同底层数据源的导入和封装
- 多租户、访问控制

Aura Visual Pipeline

Visual Pipeline

- 向导式、规范化的挖掘流程
 - 支持按算法框架的维度提供向导式、规范化的挖掘流程
- 可视化管理和监控
 - 模型训练过程监控，进度、异常、错误
 - 挖掘结果/评估结果的标准化的可视化展现
- 支持与Advanced Notebook的交互
 - 可编辑/展示开发完的模型，定义新的算法框架
- 支持与Common Data Model的交互
 - 可查看、选择以及另存为新的指标集。

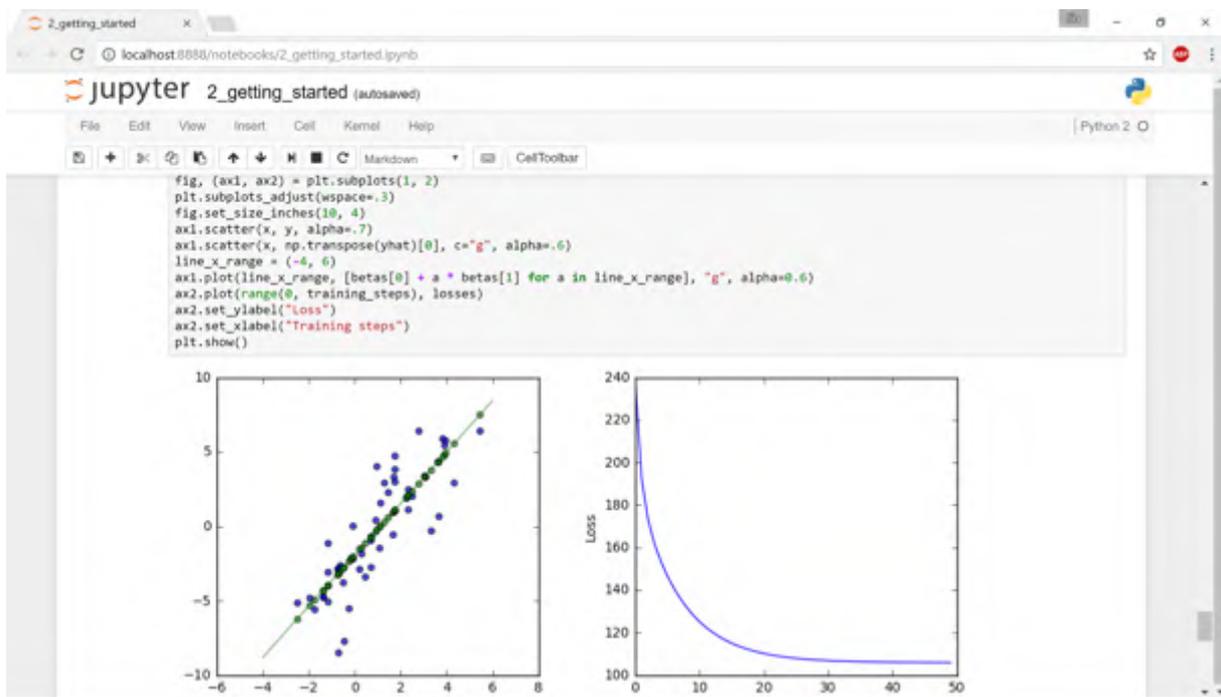


The screenshot displays the Aura Visual Pipeline interface. At the top, there are four workflow cards with completion times: 2019-04-13 12:17:30, 2019-04-12 14:13:26, 2019-04-08 17:38:41, and 2019-04-08 13:48:03. Below these are buttons for 'test 关键特征', '流量数据用户特征分析', '垃圾垃圾处理', and '最新升档策略模型分析'. A sidebar on the left lists various analysis tasks like '关键特征分析 (3)', '影响原因分析 (2)', etc. An overlay titled '选择分析框架' (Select Analysis Framework) is shown in the foreground, listing several analysis frameworks with their descriptions and icons:

- 数据智能分组**: 通过聚类分析算法，自动完成数据分组，并给出每一分组数据的特点，用户可以根据分组数据特点对各类解释和命名。
- 关键特征分析**: 分析量化事物的关键特征，并根据关键特征建立识别模型，对数据进行自动分类。
- 交叉销售分析**: 又称购物篮分析，描述在一个交易中，物品之间同时出现的规律。
- 变化趋势预测**: 站在时间的角度，把握分析对象的历史发展规律，对未来的趋势做出预见。
- 个性化推荐**: 采用协同推荐的方式，根据海量用户历史的偏好情况，预测用户潜在的偏好需求。
- 影响原因分析**: 分析事件的真正原因，量化每个因素对事件的影响程度。

Aura——Advanced Notebook

Advanced Notebook



整合开源Notebook，兼容40多种编程语言

预置丰富的计算引擎/框架及各类接口，
辅助开发开源包的版本配置

即时开发、修改，即时展现结果

支持与Visual Pipeline的交互

- 可编辑/展示Visual Pipeline中已开发完的模型
- 可将Advanced Notebook中已开发完的模型发布至Visual Pipeline

支持与Common Data Model的交互

- 提供可视化控件
- 可查看、选择以及另存为新的指标集

Aura Targeted Users

满足三类典型用户及相关需求：

用户	技术能力要求	业务能力要求	潜在需求	产品使用者归属部门	产品潜在使用人群	
					2016年	2017-2019年E
入门级用户	★	★★★★★★	✓ 现成数据应用	✓ 业务部门	75%	60%，占比递减趋势
基础级用户	★★★★	★★★★	✓ 借助简易工具及平台实现业务部门需求	<ul style="list-style-type: none"> ✓ 业务支撑部门 ✓ 亚信现场团队 	20%	30%，占比递增趋势
进阶及精通用户	★★★★★★	★★★★★★	✓ 功能强大的平台	<ul style="list-style-type: none"> ✓ 业务部门（未来） ✓ 亚信现场团队（未来） ✓ 亚信数据科学家 	5%	10%，占比递增趋势

OVERVIEW

目录

1 Retrospective and Looking Forward

大数据与人工智能技术的回顾与展望

2 AURA - AsiaInfo Machine Learning Platform

Aura 数据驱动型应用开发平台简介

3 Data-Driven Application Use Cases

数据驱动型应用示例

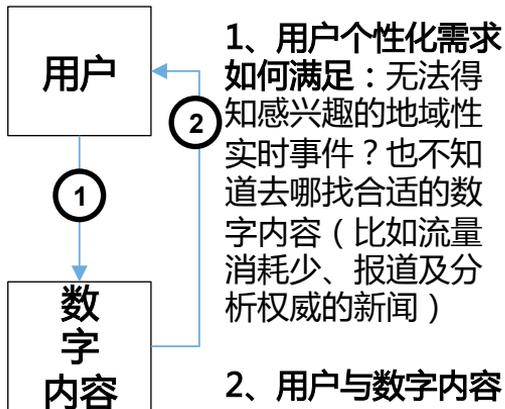
4 Discussion

问题与讨论

AI in Telco Use Case I

数字内容智慧推荐

业务痛点



1、用户个性化需求如何满足：无法得知感兴趣的地域性实时事件？也不知道去哪找合适的数字内容（比如流量消耗少、报道及分析权威的新闻）

2、用户与数字内容如何适配：向哪些用户推广即时发生的热点数字内容？

现行做法

- ✓ 针对即时发生的热点事件，依据已沉淀的用户偏好标签，粗颗粒度选取目标用户，进行短信push新闻链接
- ✓ 通过用户历史访问行为，给用户打上用户偏好标签，标签颗粒度较粗且更新不及时

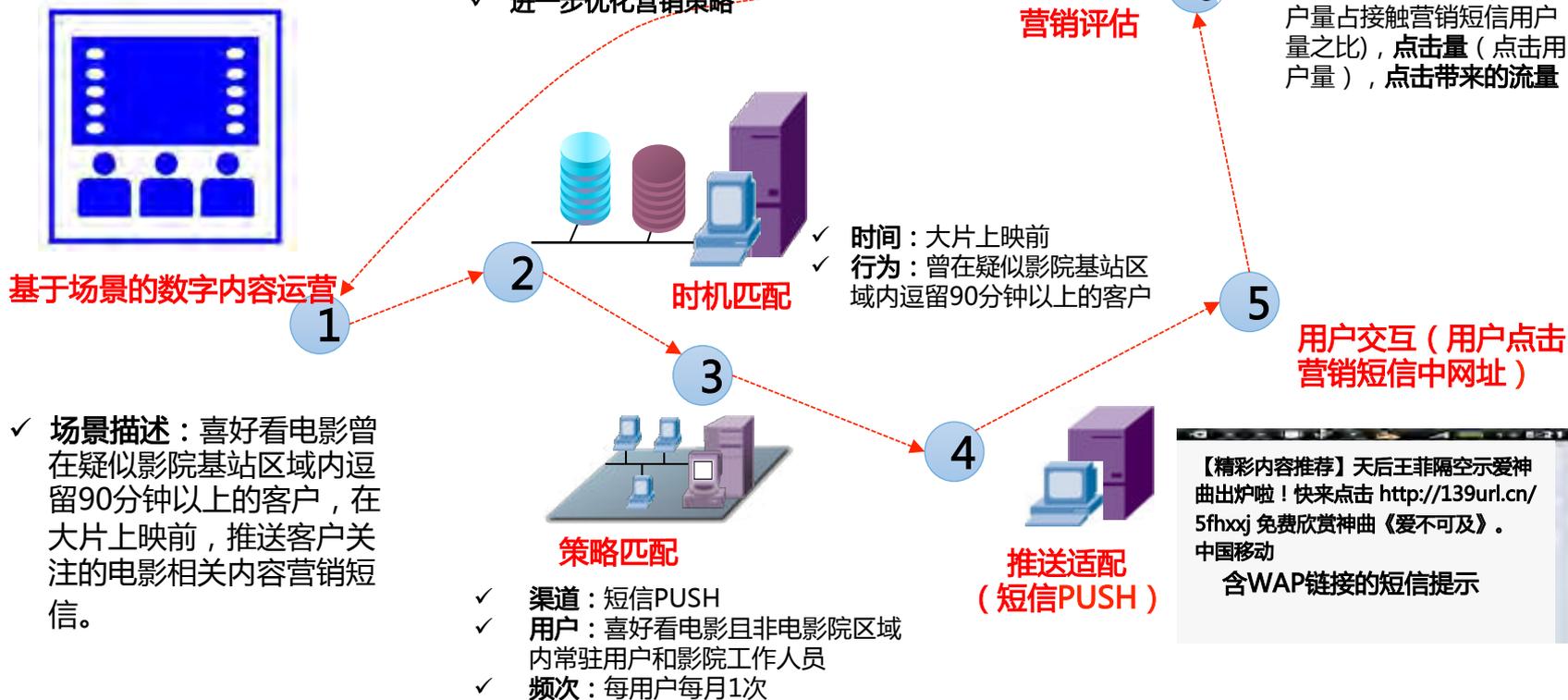
亚信做法及效果

- ✓ 快速识别捕获热点内容：按地域分为国际国内、省内、地市
- ✓ 内容源多元化：自有基地、VGOP、定向爬取（如国际/国内/地市的主流站点及APP）、新建内容资源平台
- ✓ 四种方案综合运用：1）基于数字内容分类标签和用户相似度的协同算法模型；2）集团公司标签；3）本地流量标签；4）基于业务经验的用户口径界定



AI in Telco Use Case I

数字内容智慧推荐



AI in Telco Use Case II

流量套餐/流量加油包智慧推荐

现状：根据套餐饱和度及过往月份是否有超套行为进行推荐，可推荐的用户数少、准确率低。

目标：基于全量指标对比和先进的特征选择、决策树等算法，找到影响超套行为的关键因素。即通过数据驱动，找出各档套餐/流量加油包的目标用户筛选规则，不再完全依靠业务经验。

指标选取

传统模式

- 过往月份是否有超套行为
- 套餐饱和度

建议模式

- 全量指标对比，确保找到影响超套行为的关键因素

用户消费特征

- 上月指标简单组合等于用户次月消费特征
- 上月指标预测用户次月消费特征

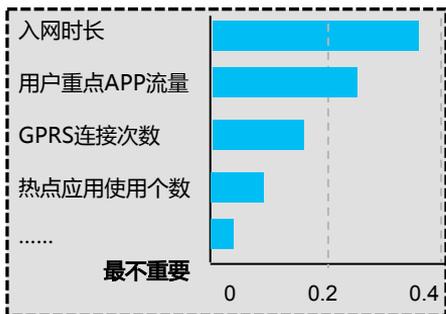
工具方法

- 业务经验
- 业务经验+机器学习

效果评估

- 工单流程
- 实时评估

流量超套关键影响因素分析



自动给出各档套餐/流量加油包筛选规则

```

age_age_bill == 2.789 [模式 0]
age_1_cmsat_count == 95 [模式 0]
age_cmsat_count == 95 [模式 0]
age_cmsat_count == 35 [模式 0]
age_cmsat_count == 32 [模式 0]
age_cmsat_count == 32 [模式 0]
intel_flag == 1 [模式 0]
brand_id == 3 [模式 0]
ms_duration == 78 [模式 0]
age_cmsat_count == 47 [模式 0]
intel_flag == 1 [模式 0]
intel_flag == 1 [模式 0]
ms_duration == 47 [模式 0]
age_age_bill == 2.013 [模式 0]
age_age_bill == 2.013 [模式 0]
ms_duration == 47 [模式 0]
    
```

自动给出各档套餐/流量加油包目标用户

建议升档套餐名称	40元手机上网流量包	50元手机上网流量包	100元手机上网流量包
商丘	3213人	312人	42人	10人
郑州	11435人	927人	249人	106人
安阳	3906人	226人	52人	15人
新乡	4343人	308人	57人	23人
许昌	2616人	191人	41人	11人
.....	2583人	299人	39人	24人

AI in Telco Use Case II

流量套餐/流量加油包智慧推荐

场景：
基于模型+时
机触发

月初生成流量套餐
潜在升档用户明细



场景描述：月初个性化实时推荐流量套餐平台已得到一批次月潜在各档流量套餐升档用户明细。

1

✓ 进一步优化营销策略

营销评估

6

✓ 活动参与率；流量套餐订购率；模型预测准确率

2

时机匹配

✓ 时间：非工作时间
✓ 行为：用户通过手机终端浏览器浏览网页且该网页页面产生流量超过用户剩余可用流量30%



用户订购流量套餐

3

策略匹配

✓ 渠道：短信PUSH

4

推送适配
(短信PUSH)

尊敬的用户您好，您当前使用流浪网页消耗流量过多，您本月剩余可用流量不足，建议您订购X元流量套餐，现在订购还可参加“达量1G即赠1G”促销

短信提示

• 注：非工作时间：周一至周五（18点~20点）、周六、周日（8点~20点）

AI in Telco Use Case III

终端合约智能推荐

背景说明

终端引领一直是运营商很重要的用户发展手段，终端合约可以帮助运营商吸引客户、留住客户。因此通过增加对终端合约的用户信用等级评估模型和潜在终端合约用户精确识别模型，可以帮助运营商找到正确的终端合约潜在订购和续约用户，并向其进行准实时的PUSH营销，以此完成终端合约推荐，增加用户粘性，提升客户保有和客户价值。

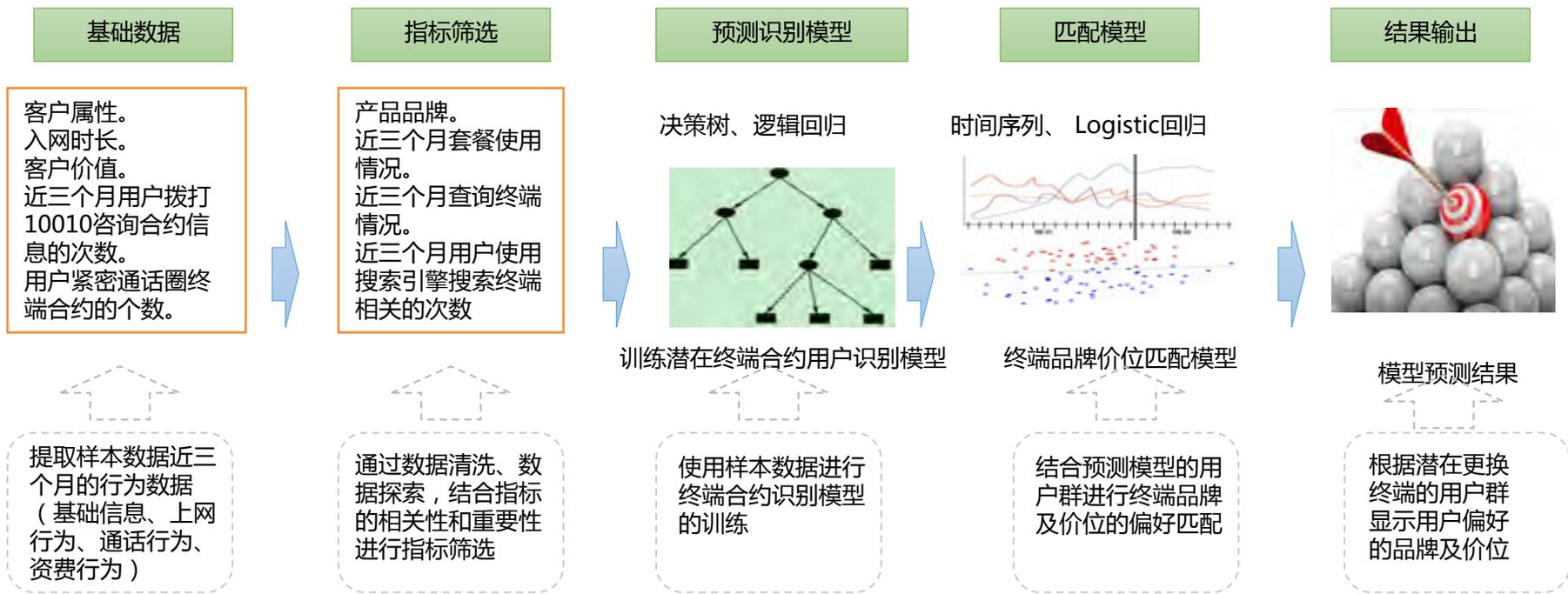
应用目标

结合标签体系的基础上，分析用户互联网行为，预测用户换机行为，挖掘终端类合约计划潜在订购和续约用户，并向其进行准精准营销，增加用户粘性，提升客户保有和客户价值。

AI in Telco Use Case III

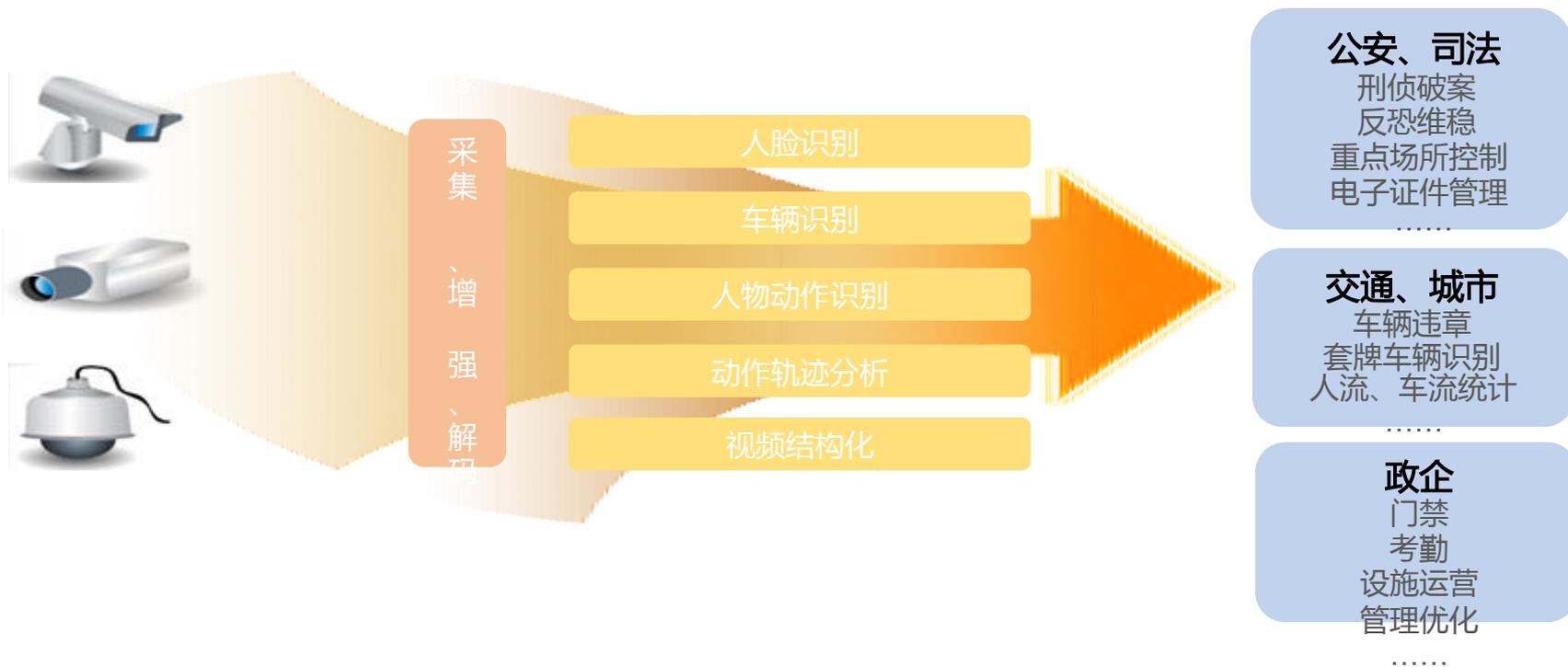
终端合约智能推荐

分析思路：本次终端合约用户挖掘模型是基于全网的订购合约前的上网行为、通话行为等原子属性的分析终端合约用户的关键特征，结合关键特征预测潜在更换终端的用户群。再根据终端合约包含的原子属性，将潜在更换终端用户进行终端匹配，为用户匹配用户可能偏好的终端品牌。

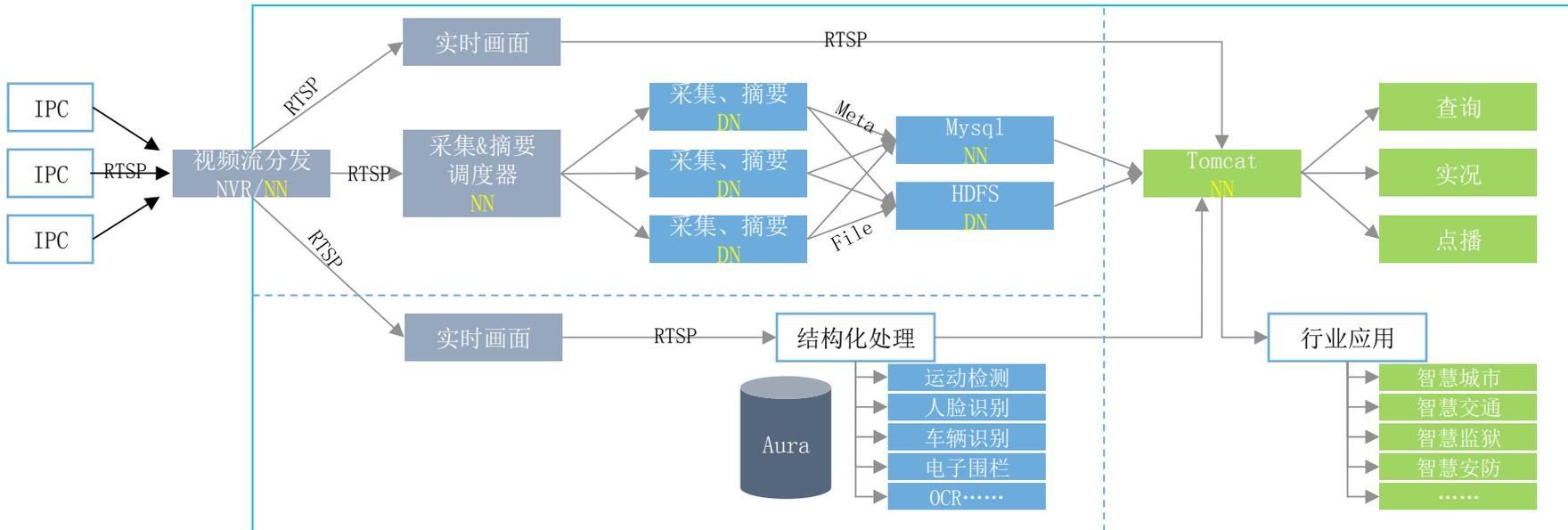


基于Aura的深度学习应用

基于Aura平台，我们实现视频图像大数据的智能分析和人脸识别，能快速发现、定位和识别特定目标，并进一步与位置运营和智慧交通等大数据应用结合，可广泛应用于安防监控、刑侦和反恐等领域；



应用系统架构



场景应用



- 身份数据
- 人流数据
- 习惯数据
- 热门门店数据
- 设施使用情况数据



- 消费行为数据
- 网络电子足迹
- 位置数据

.....



- 客户定位
- 智能推送
- 优化布局, 运营

.....



- 人流数据
- 个体行为数据
- 车流数据
- 车型数据
- 拥堵数据

.....



- 位置数据
- 交管数据
- 车辆数据
- 驾驶员数据

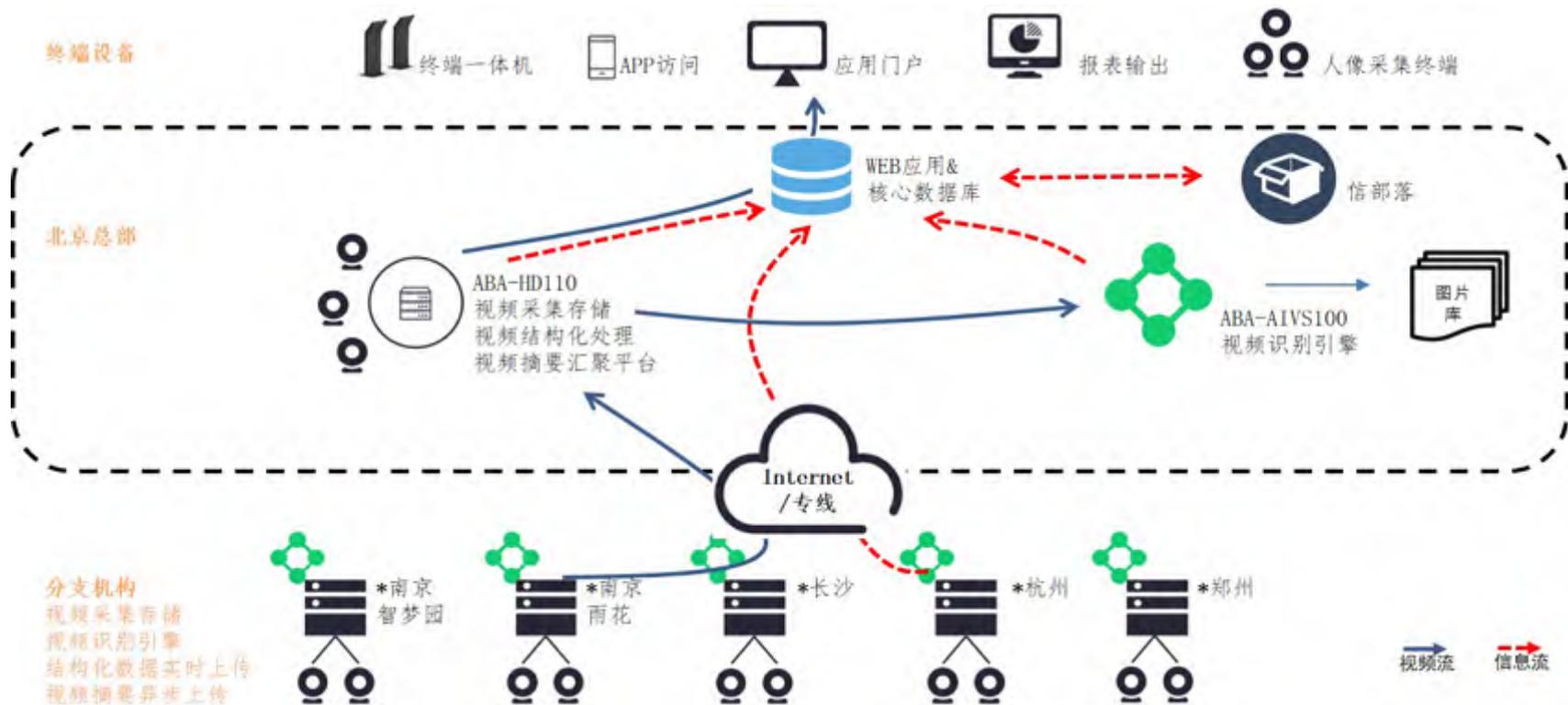
.....



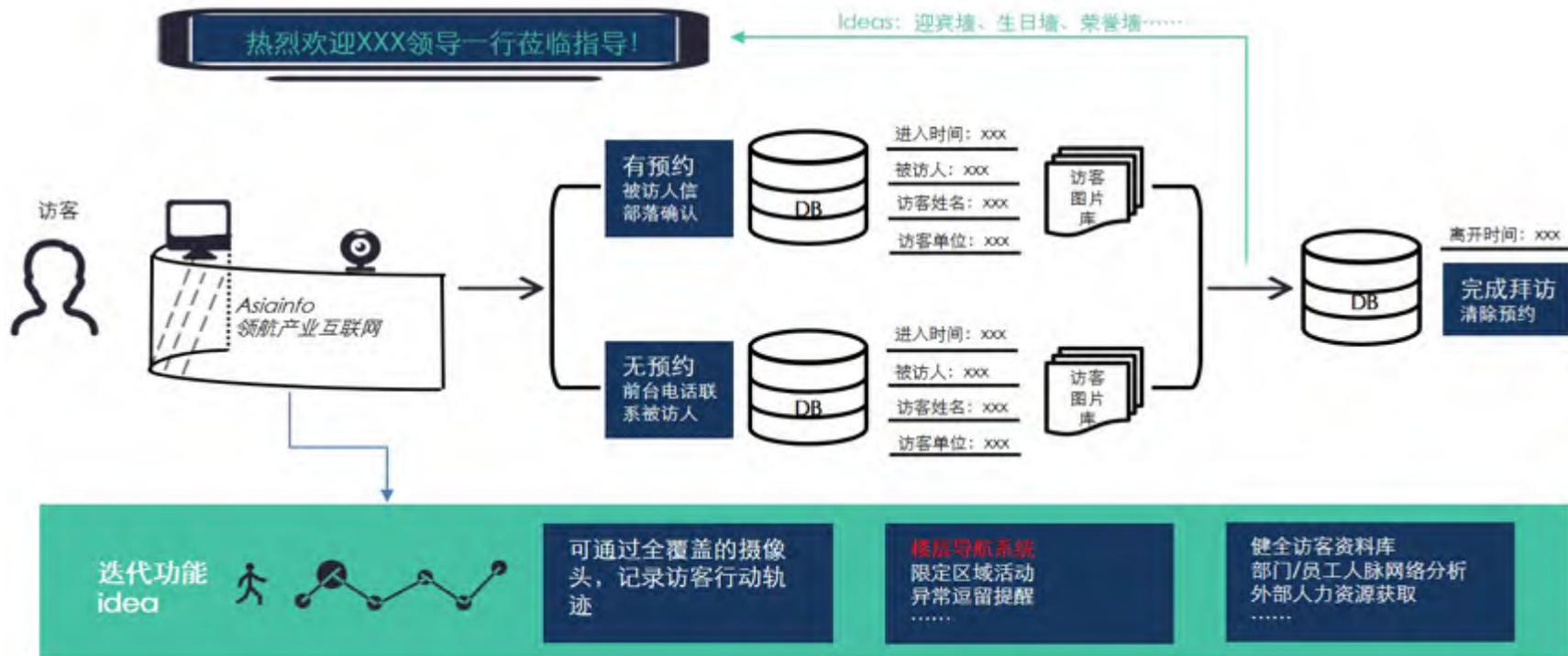
- 智能交通
- 人/车流疏导
- 实时预警
- 智能服务推送

.....

亚信“大脑”



访客系统



行政管理

员工出入记录					
姓名	工号	部门	生日	进门时间	出门时间
张三	16789	CIT-XXX-XXX	1987-XX-XX	2016/9/12 8:35	2016/9/12 18:35
李四	26789	CMC-XXX-XXX	1988-XX-XX	2016/9/12 8:45	2016/9/12 18:45
王五	66789	BDX-XXX-XXX	1989-XX-XX	2016/9/12 8:55	2016/9/12 19:55
赵六	76789	CUC-XXX-XXX	1990-XX-XX	2016/9/12 9:15	2016/9/12 20:15
钱七	36789	CTC-XXX-XXX	1991-XX-XX	2016/9/12 8:35	2016/9/12 18:35
孙八	56789	CIT-XXX-XXX	1992-XX-XX	2016/9/12 8:45	2016/9/12 18:45
周九	86789	BDX-XXX-XXX	1987-XX-XX	2016/9/12 8:55	2016/9/12 19:55
吴十	96789	CUC-XXX-XXX	1988-XX-XX	2016/9/12 9:15	2016/9/12 20:15
小明	46789	CMC-XXX-XXX	1989-XX-XX	2016/9/12 8:35	2016/9/12 18:35
小红	39876	CTC-XXX-XXX	1990-XX-XX	2016/9/12 8:45	2016/9/12 18:45
小芳	29876	CIT-XXX-XXX	1991-XX-XX	2016/9/12 8:55	2016/9/12 19:55
小军	19876	BDX-XXX-XXX	1992-XX-XX	2016/9/12 9:15	2016/9/12 20:15
.....



考勤时间统计



员工自动考勤



部门出勤统计



工位布局优化



工作时长统计



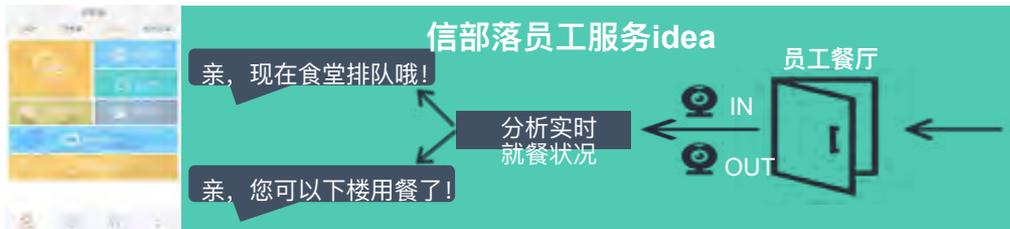
加班员工关怀



楼内人数统计



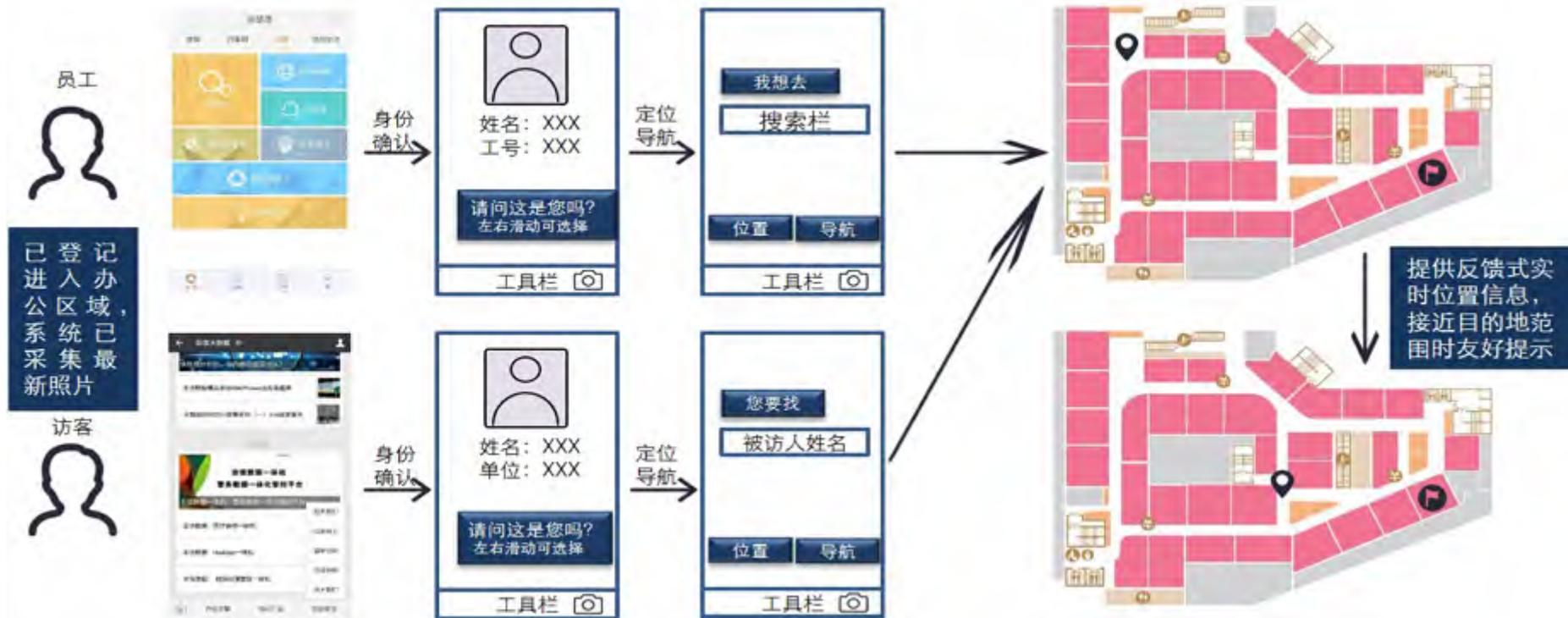
餐食合理定量



安全管控



楼内导航



OVERVIEW

目录

1 Retrospective and Looking Forward

大数据与人工智能技术的回顾与展望

2 AURA - AsiaInfo Machine Learning Platform

Aura 数据驱动型应用开发平台简介

3 Data-Driven Application Use Cases

数据驱动型应用示例

4 Discussion

问题与讨论

THANK YOU

By 亚信数据