

「阿里技术参考图册」 「算法篇」

前言

《阿里技术详解图册》，分为研发篇、算法篇两册，将为你清晰呈现阿里业务生态的全貌：文化娱乐（优酷、土豆等）、核心电商业务（天猫、淘宝、村淘等）、本地生活（高德地图、盒马等）、支付 & 金融服务（蚂蚁金服）、智慧物流（菜鸟）、市场营销（阿里妈妈）、云计算（阿里云）等。

同时，你更能高屋建瓴、深入了解每个业务背后的技术支撑体系：高性能基础设施如何支撑商业系统？集群调度系统高效怎样提升资源效率？新一代分布式存储引擎、数据库技术、分布式中间件有哪些创新之处？同时，新一代计算平台打造的世界级计算能力、达摩院机器智能实验室引领的技术和产业创新、新零售开启的 DT 时代消费新体验，也将一一展现在你的面前。

如果你希望加入阿里搞技术，却面对不同的部门、成百上千的技术岗位一头雾水，这套图册将为你拨开迷雾，清晰呈现每个事业部的主攻方向、技术创新及突破、发展路径等。特别是对于 JAVA 工程师、C/C++/ 工程师、数据研发工程师、数据分析师、算法类工程师，此书具有巨大的参考价值。

如果你希望了解业界最新技术趋势，来自阿里人工智能实验室、天猫、淘宝、菜鸟等部门技术负责人，同样带来前沿的技术解析，揭秘阿里在各个领域的技术积累及发展方向。

受限于篇幅原因，此书未对各个技术领域进行深入讲解。如有兴趣，可关注阿里技术官方公众号，查阅 300 篇 + 阿里技术精华干货（持续更新中）。

希望这套图册，能为你打开一扇窗户，去看更大的世界；成为一个小支点，帮你撬动更大的进步。

让我们一起遇见更好的自己。



微信扫描二维码
关注阿里技术



钉钉扫描二维码，或搜索群号
21764737，加入读者交流群

目录

走进阿里技术

人工智能实验室简介

人工智能实验室研究员 永川

走进阿里妈妈技术

阿里妈妈 靖世

从人工智能到机器智能

达摩院机器智能实验室 智捷

阿里云：为了无法计算的价值

阿里云资深专家 蔡华

走进淘宝技术部算法世界

商业机器智能部资深算法 永叔

天猫技术：电商“最强大脑”

营销平台资深算法专家 志昭

搜索事业部：AI@ 搜索和推荐

资深算法专家 三桐

新零售供应链平台事业部：智慧供应链

资深技术专家 粤谦

走进蚂蚁金服技术

蚂蚁金服资深技术专家 南笑

菜鸟网络智慧物流：算法应用的新蓝海

菜鸟网络人工智能部资深算法专家 元享

阿里巴巴跨境 B2B 中的算法技术

国际技术事业部 资深算法专家 守清

走进阿里技术



繁荣的业务生态

文化娱乐



核心电商业务



本地生活



支付&金融服务 PAYMENT & FINANCIAL SERVICES



智慧物流 LOGISTICS



市场营销 MARKETING SERVICES & DATA MANAGEMENT



云计算 CLOUD COMPUTING

超大规模的企业级应用

5470亿_{美元}

2017财年商品交易额GMV (美元)
直逼全球前20经济体

4.54亿

年度活跃买家
(2017年度)

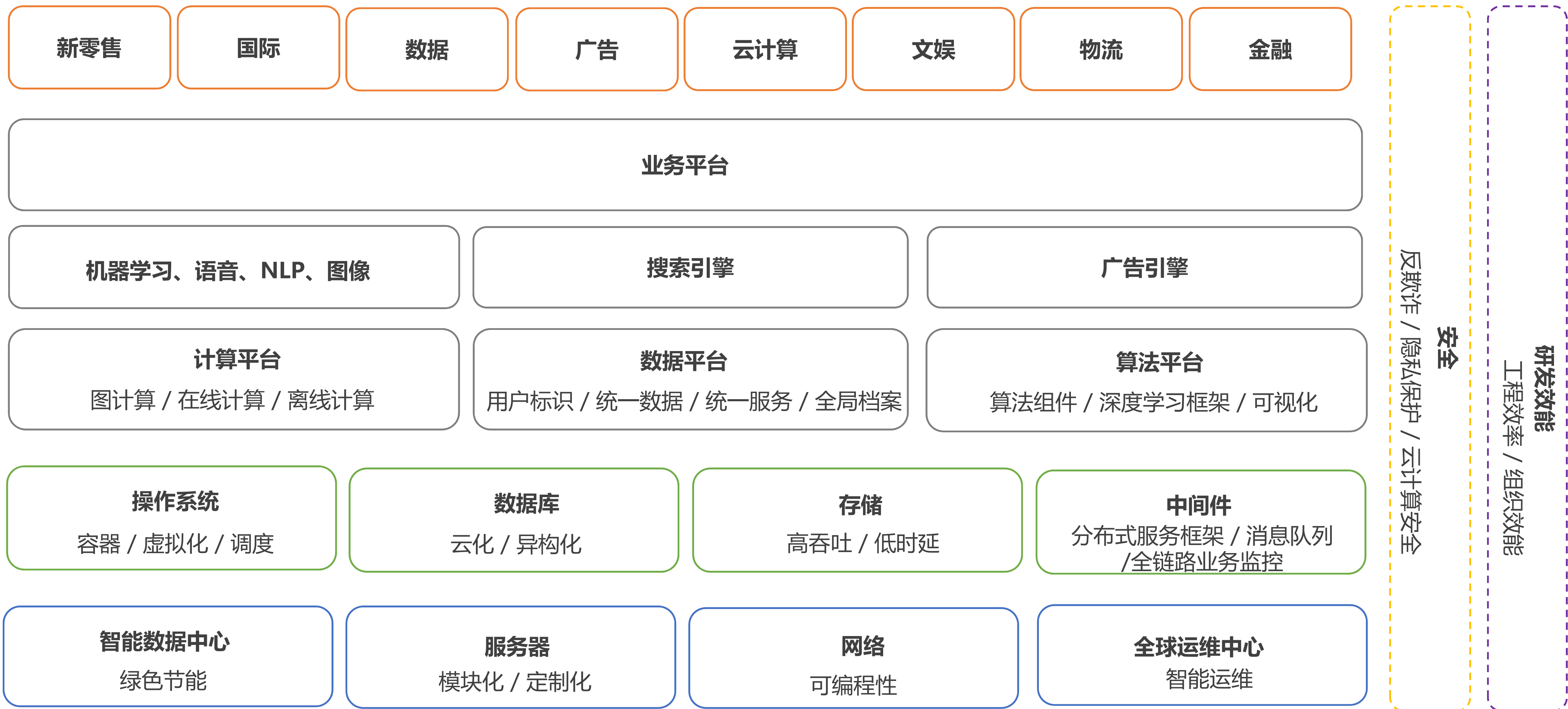
5.07亿

移动端月活跃用户MAU
(2017年3月)

EB 级数据处理

参考：阿里巴巴2017财年 (2016年4月1日~2017年3月31日) 财报中的**中国零售平台**数据。

大中台、小前台的技术组织架构



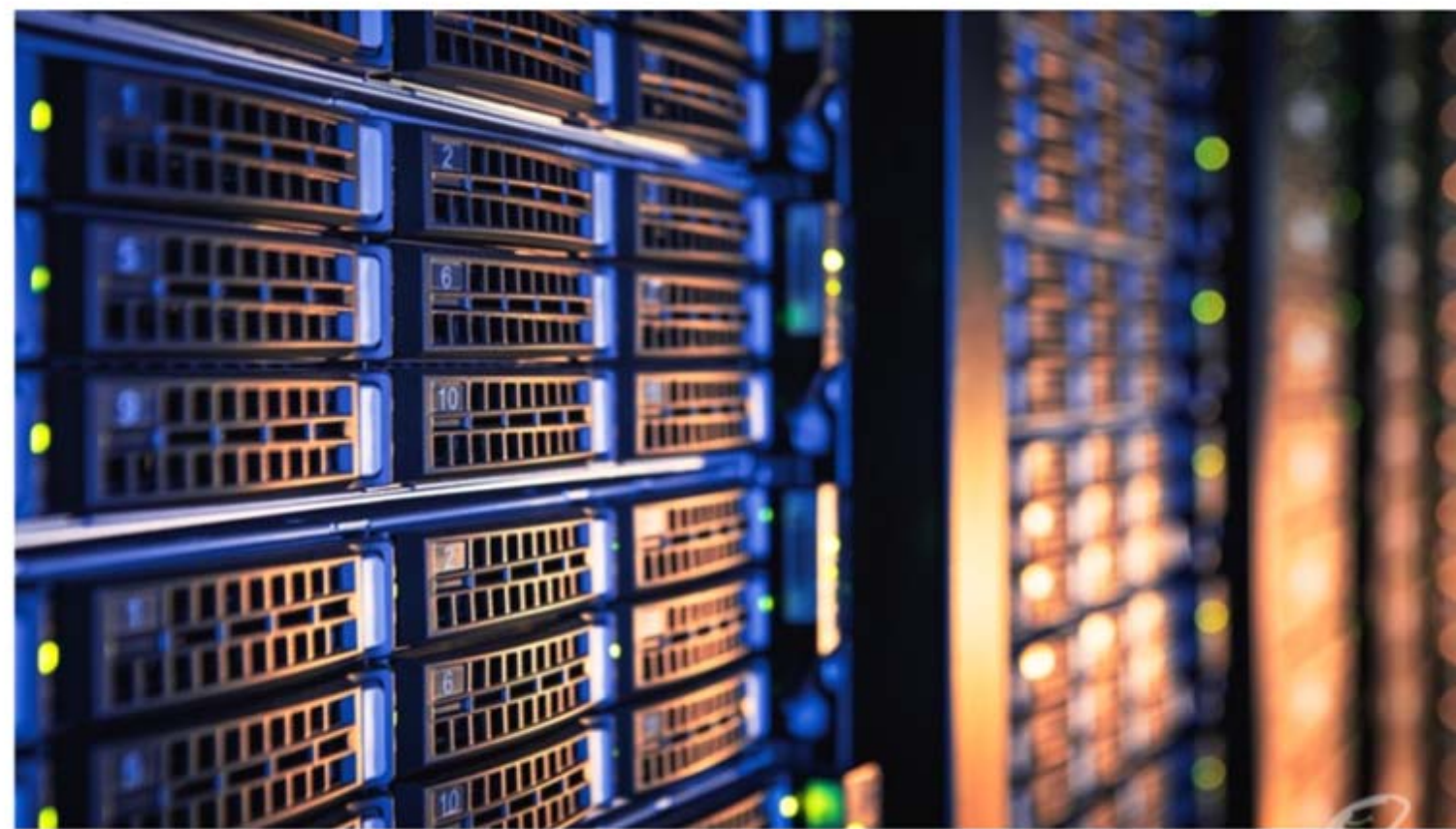
高性能基础设施支撑商业系统



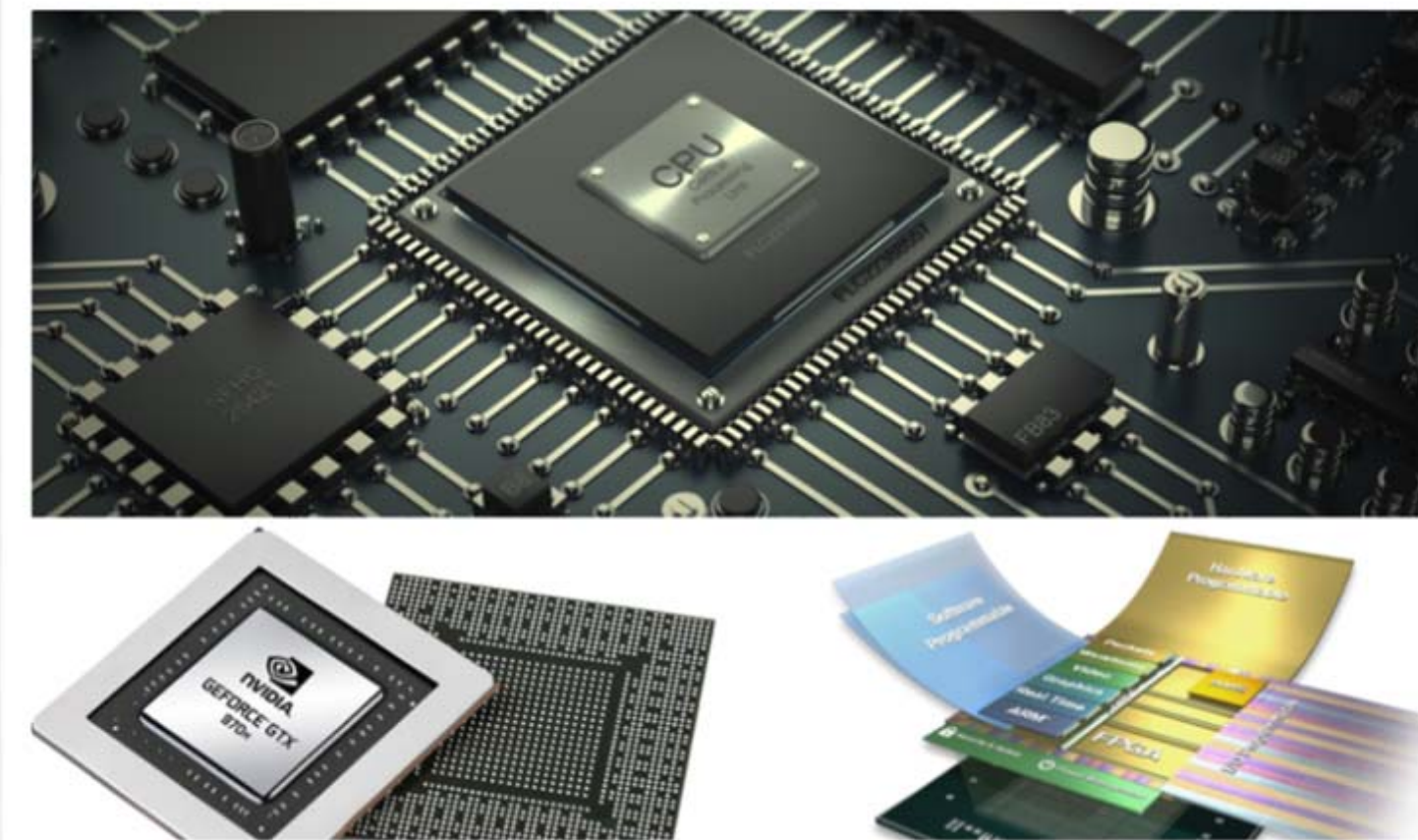
绿色数据中心：互联网行业全球首个浸没液冷服务器集群，PUE逼近理论极限值1.0



自研网络：大规模、高性能、低延迟，从自动化到智能化，改善全球用户体验



定制化芯片与服务器：芯片、存储设备、服务器的全面自研



异构计算：国内首个FPGA资源池，支持大型计算任务的并发处理

集群调度系统高效提升资源效率

10% -> 40%

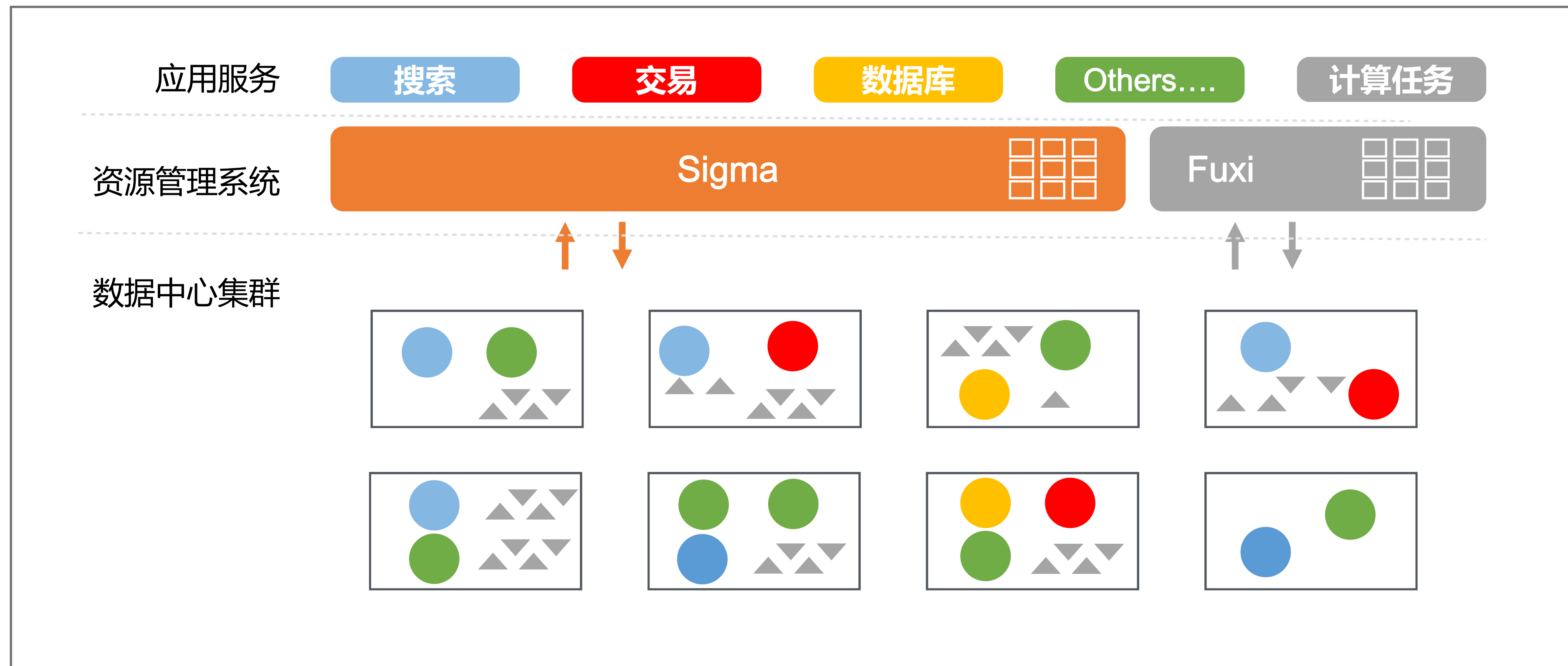
CPU平均利用率业内低于10%

万台

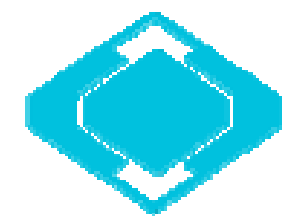
混合部署集群规模

30%

节省服务器资源



面向未来场景的分布式存储引擎



块存储



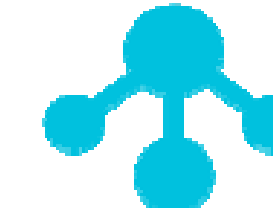
文件存储



对象存储



归档存储



表格存储

99.99999999%

数据可靠性

10PB

单个文件系统
弹性付费

48.8TB

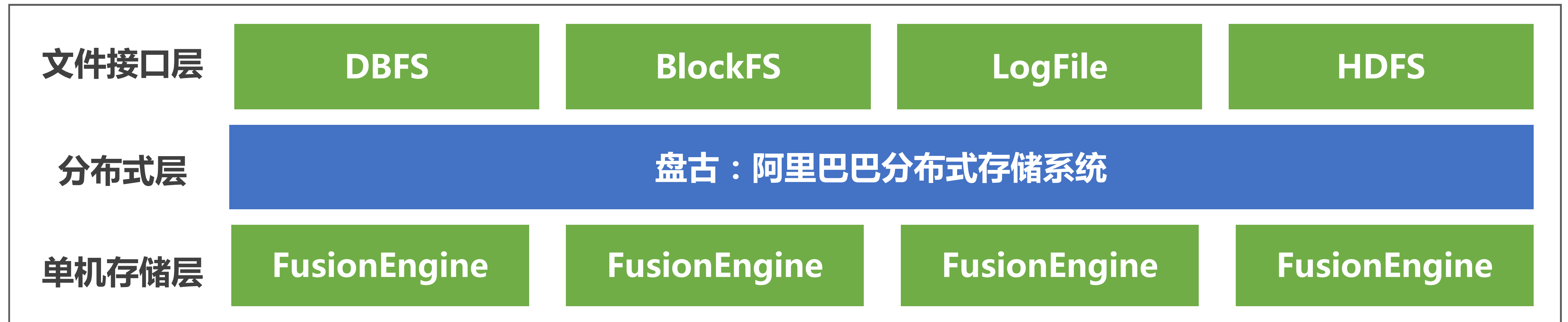
单文件
丰富的数据处理生态

78%

最高成本降低
冷热数据自动管理
完全兼容OSS API

毫秒级

延迟响应
支持单表千万级并发



全线数据库技术与产品

X-DB

世界上最快、成本最低的OLTP数据库

- 原生设计的分布式数据库
- 全球化部署，大于**5个9**的可用率
- 数据强一致
- **10倍**MySQL性能，**百万TPS**
- **1/10** MySQL存储成本
- 全面兼容MySQL生态体系

AnalyticDB

全球领先的超大规模极速OLAP云数据库

- **千级**节点数、**百PB级**数据量
- **十万级**QPS、每秒**千万条**写入
- **GPU 异构**计算加速，10倍性能提升
- 毫秒级**非结构化数据**检索与融合分析
- 基于AI的智能化优化、自服务
- 兼容MySQL生态、分析函数大幅增强

POLARDB

国内首个Cloud Native商用关系数据库

- 单实例**百万QPS**
- 3副本**微秒级**持久化
- **100%**兼容MySQL和PostgreSQL
- 读写性能按需阶跃
- 百TB容量**分钟级**备份
- 国内首款数据按量计费



新一代数据库技术-软硬件结合



新一代数据库技术-智能数据库

高性能分布式中间件助力企业互联网数字化转型

技术领先

1. 分布式服务框架支持机器集群规模百万，故障自动识别率>95%，自愈时延<3秒
2. 世界首次实现了基于在线交易系统的全链路压力仿真
3. 亿级消息洪峰下99.996%消息时延小于3ms

技术开源

1. Aliware开源技术**10**项，被国内众多企业采用
2. 2017年9月分布式消息中间件RocketMQ 成为阿里巴巴首个Apache顶级开源项目
3. Aliware挑战赛：500高校、300企业、3000队伍



2015年中国石化采用Aliware实现3个月快速上线易派客电商交易平台，累计交易额突破1000亿
易派客列为金砖国家2017年八项核心成果之一
成为服务金砖国家的工业品交易电子商务平台

新一代计算平台打造世界级计算能力

99% / 95%

统一大数据平台

承担阿里 99% 的数据存储
与95% 的计算任务

60K+ / 10+

超大规模，跨域容错

60000+ 服务器，10+ 集群
跨DC调度容灾能力

4M+ / 300PB+

快速增长

作业量400万/日，增长 60%
处理数据量超300PB/日，增长80%



	TPC-BB 此前最佳成绩	阿里巴巴 MaxCompute 2.0 + PAI	新突破
测试最大规模	10TB	100TB	首次将规模拓展至100TB
最佳性能	1491.23QPM	7830QPM	首个达到7000分的引擎
最高性价比	\$589.91 /QPM	\$ 354.7 / QPM	首个基于公有云服务的 benchmark

机器智能引领技术和产业创新



语音智能



视觉智能



自然语言
理解



决策智能



深度学习

天猫精灵

100万家庭语音交互入口，“我在，你说”



拍立淘

以图搜图，支持搜索千亿图像



智能地铁

语音购票+刷脸支付
刷脸进站+客流分析



互联网汽车

斑马智行，搭载上汽
荣威RX5，40W+辆



智能客服

分担双11当天95%的业务咨询量



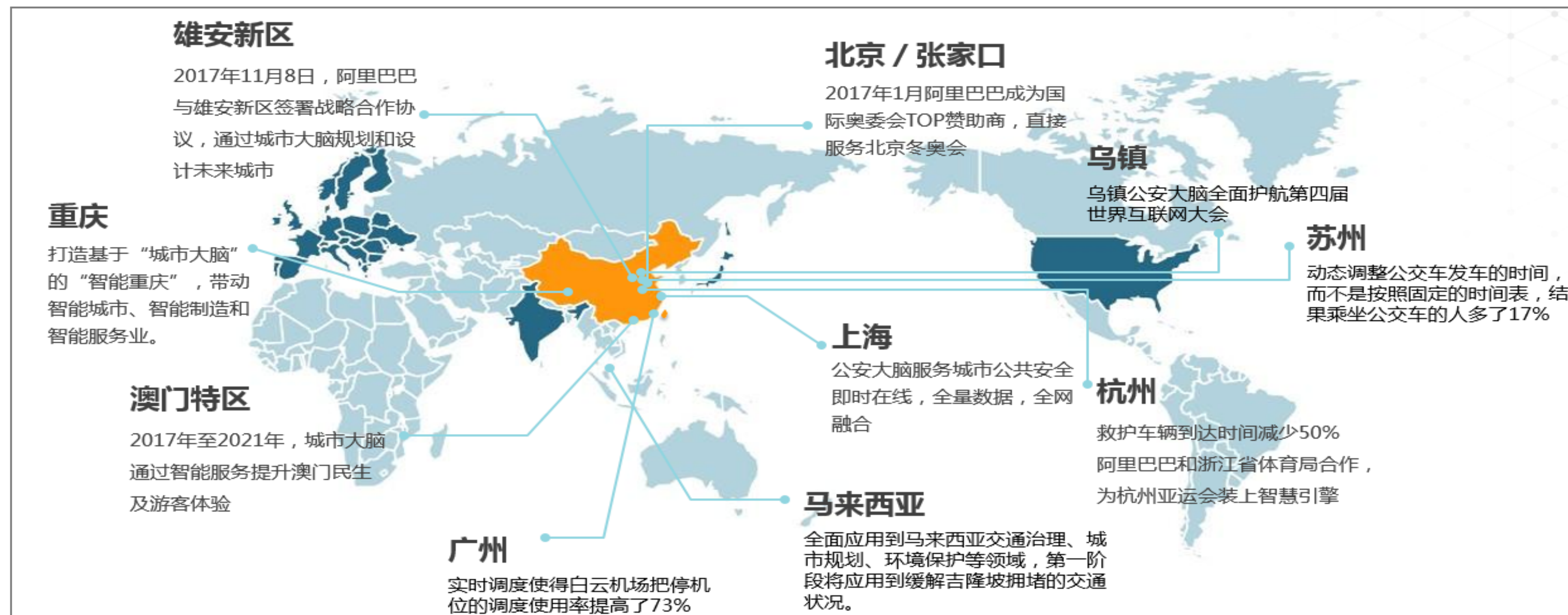
阿里小蜜

城市大脑让城市自我思考和自我治理

城市大脑：将散落在各个角落的数据汇聚到一起，使用**云计算、大数据和人工智能技术**，让城市的各个“器官”协同工作，变成一个能够自我调节、与人类良性互动的有机体。

1. 信号灯在线优化配时，平均通行速度提升**15%**，交通事件实时感知正确率达到**95%**
2. 特种优先通行，救护车到达时间节省**50%**
3. 已在全球多个城市开展试点，未来3年至少服务100座城市

2017年11月城市大脑入选首批国家人工智能开放创新平台



云边端一体化让万物连接、在线、智能

端

2017年10月AliOS Things开源，
对接16家芯片厂商和52家模组，
装机设备超560万台

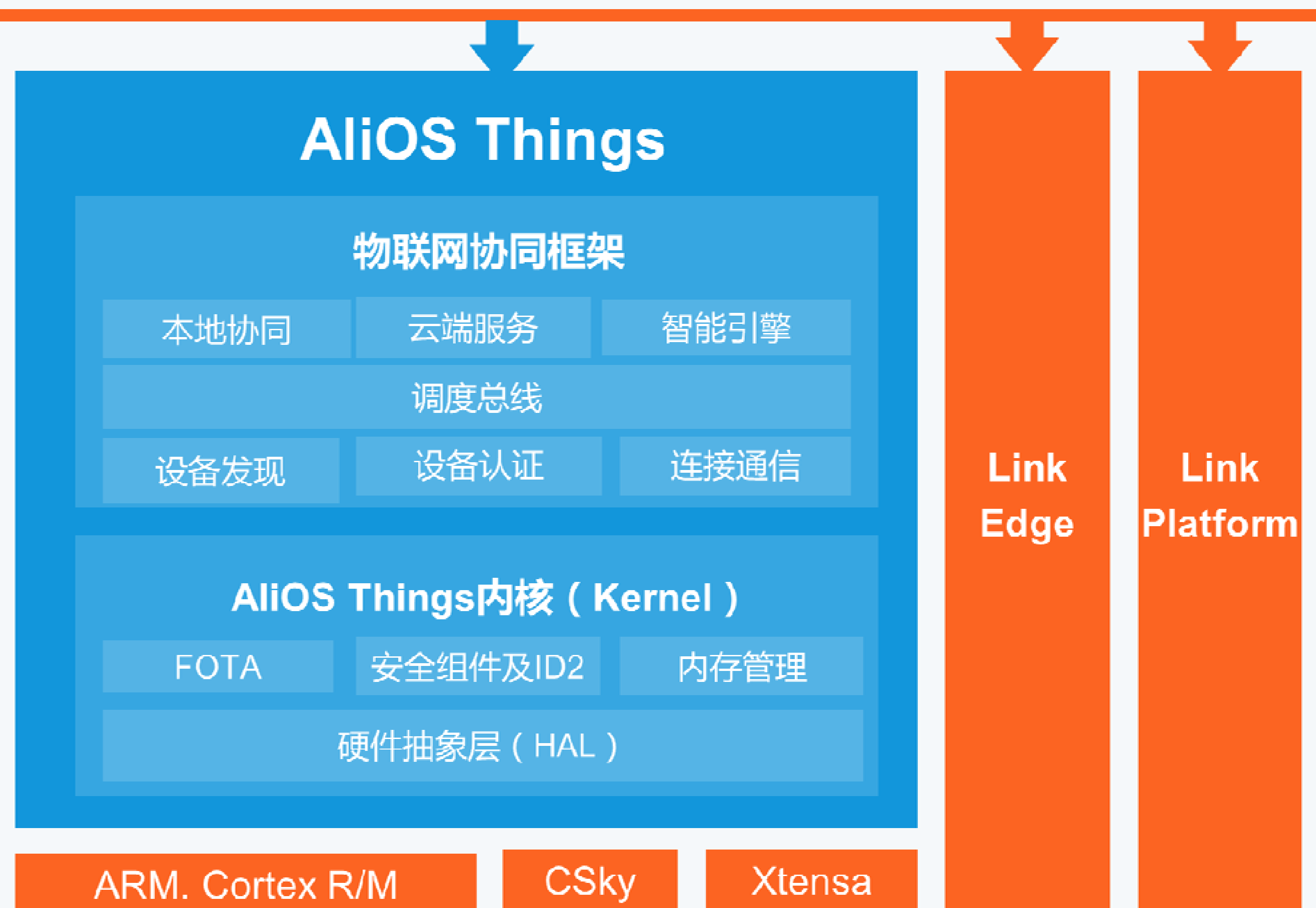
边

LinkEdge成功应用雄安数字酒店的设备管理、人脸识别等本地高阶应用

云

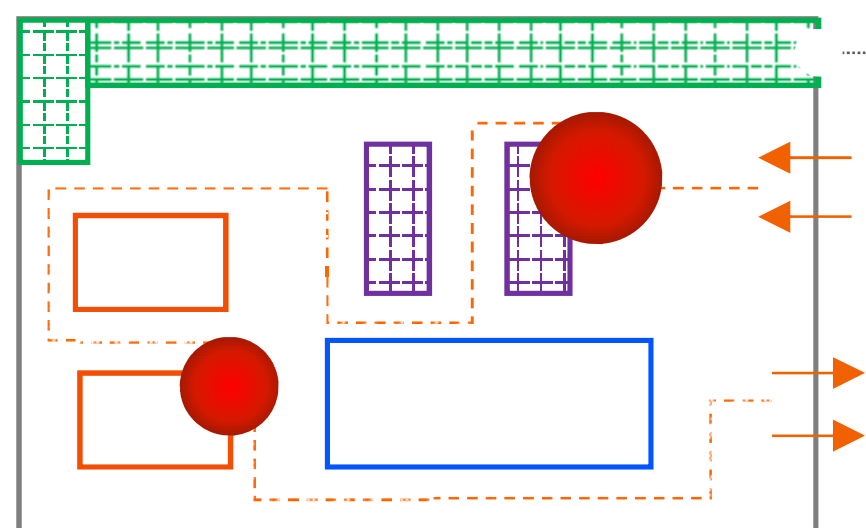
具备套件、边缘计算等能力

物联网云边端一体化计算协同框架

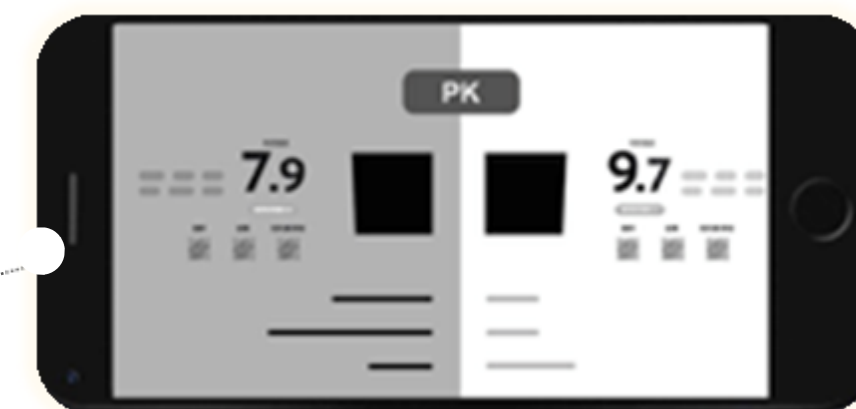
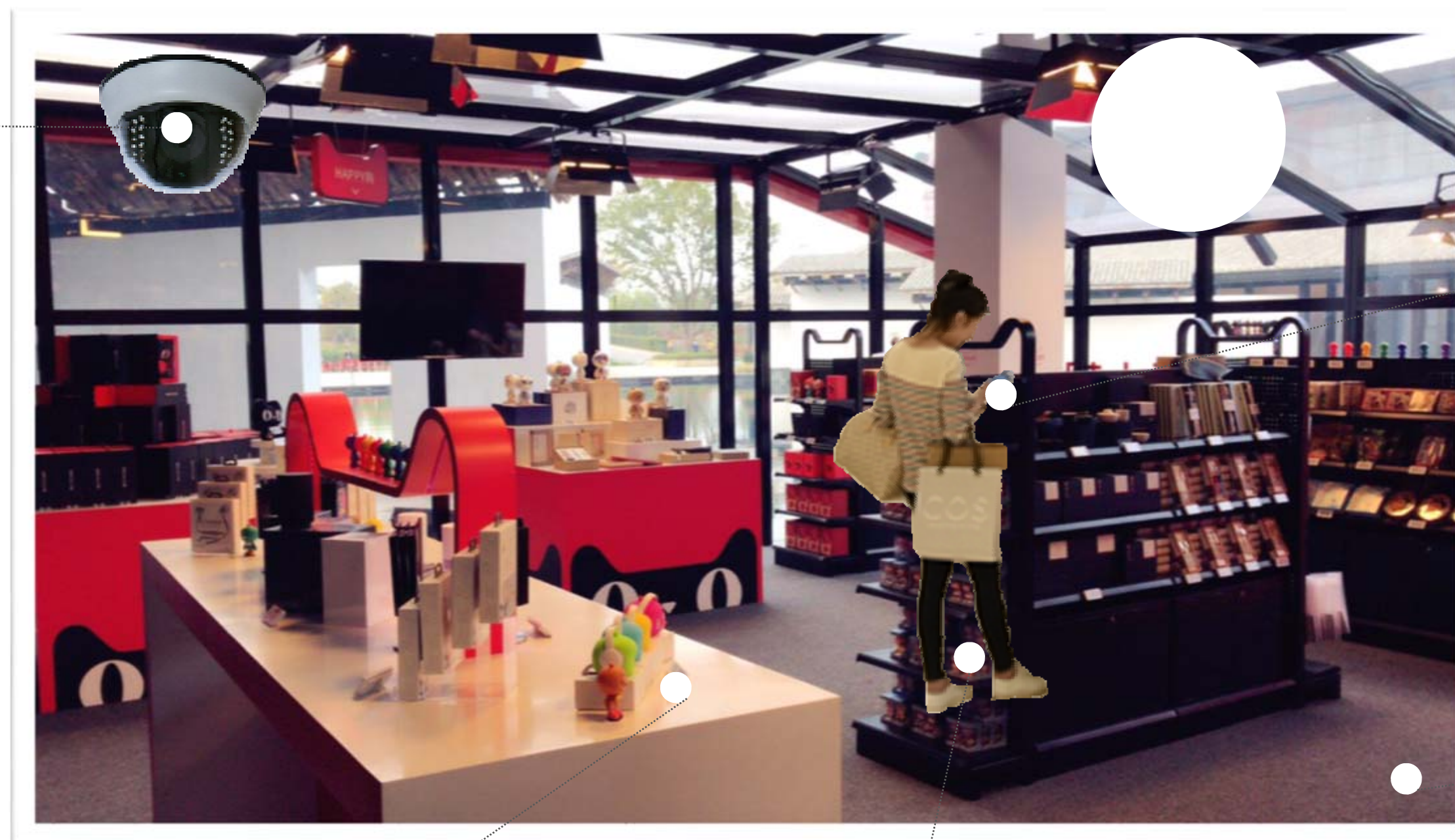


新零售开启DT时代消费新体验

线下门店数字化重构，打造品牌、门店、消费者共建、共赢生态



全域追踪，实时分析，
千人千面



线上线下数据打通



视觉识别，无人结算

电子价签
信息实时更新



货场人重构，定制化供货，
智能增值业务投放

达摩院 以科技，创新世界

未来产业驱动

城市大脑

互联网汽车

物联网

金融科技

技术研究方向

语音智能

计算视觉

自然语言理解

决策智能

无人驾驶系统

无人物流系统

异构计算平台

大规模数据处理平台

量子计算

生物识别

区块链

金融智能

全球研究中心

北京

杭州

深圳

新加坡

以色列

西雅图

硅谷

俄罗斯

学术合作网络链接全球智力资源



人工智能实验室简介

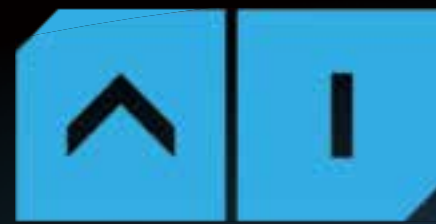
人工智能实验室研究员 永川

人工智能实验室

研发State-of-the-art人工智能技术，开发可落地的人工智能产品



数据
Big Data



Alibaba A.I. Labs
阿里巴巴人工智能实验室



预测
Forecast



深度学习
Deep learning



语音
Voice



图像
Image

我们来自



技术积累

实验室研发人员发表论文超过200篇 ★

★ Google Scholar引用次数过万

CVPR 2018发表论文9篇 ★

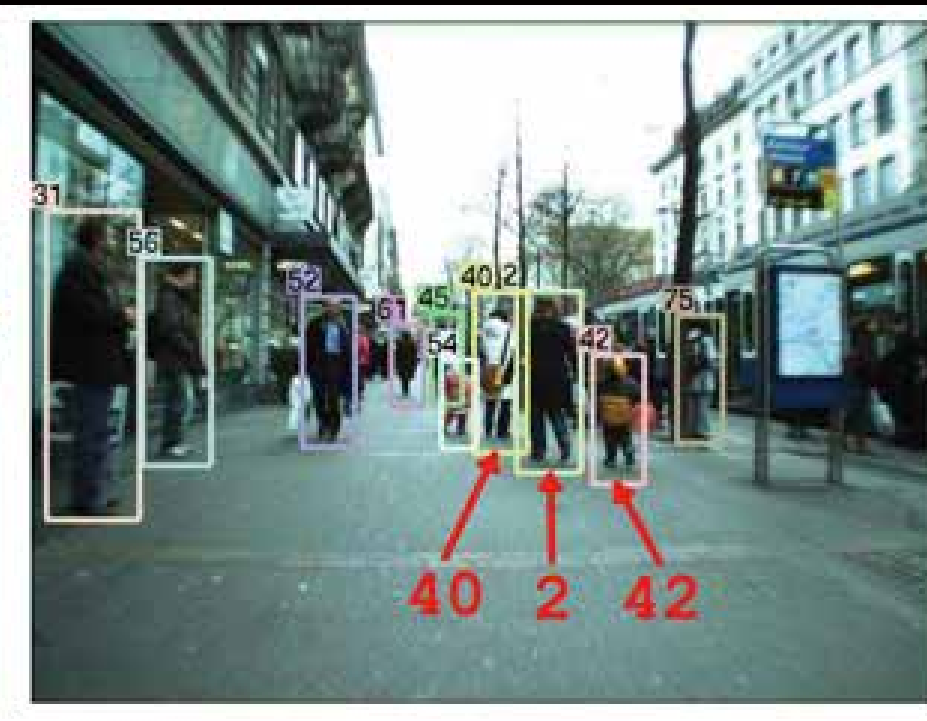
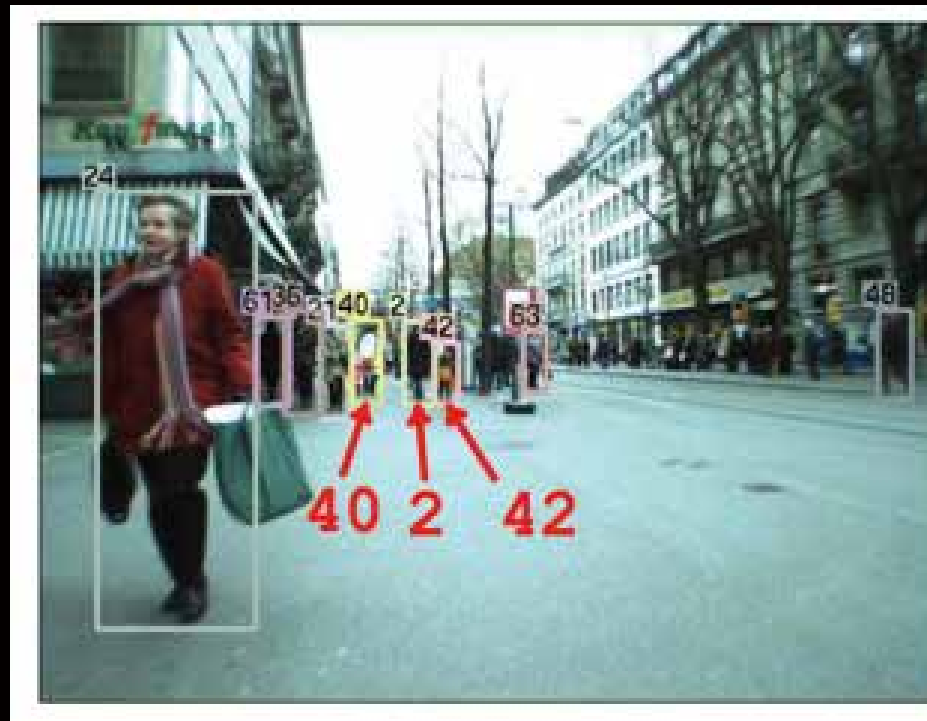
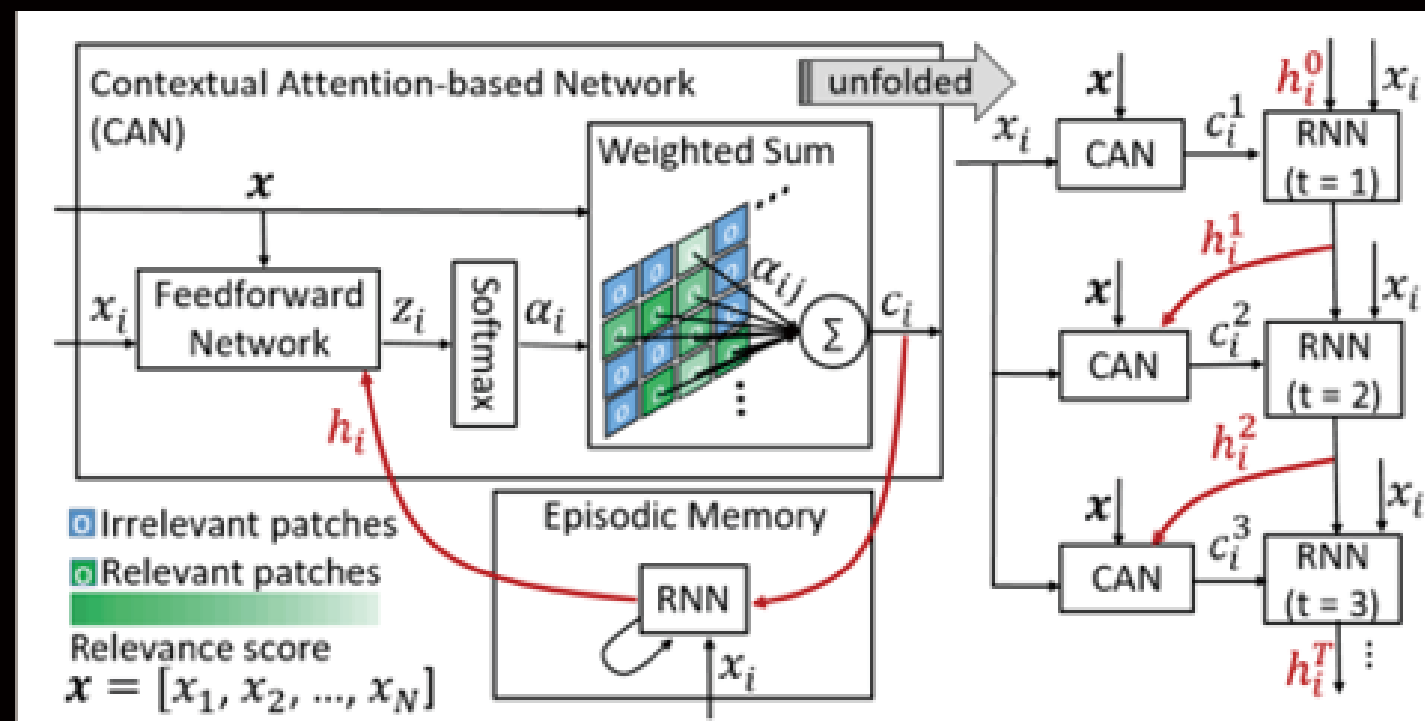
★ AAAI 2018发表论文3篇

多次获得最佳论文奖 ★

★ 多次获得国际竞赛冠军

数篇经典论文对领域产生了重大影响 ★

★ 有国家千人等知名专家



愿景：打造下一代人机交互方式和自动驾驶

使命：让机器不仅拥有计算能力，而且还拥有知识，
推理能力，以及行动力

下一代人机交互

KEYBOARD



键盘
DOS系统

MOUSE



鼠标+键盘
Windows

TOUCHSCREEN



触屏
iPhone

VOICE, etc.



用嘴,眼睛,耳
AI 时代

下一代人和机器的交互方式，应该像人和人交互一样更加接近自然

阿里巴巴推出
第一款消费级人工智能产品

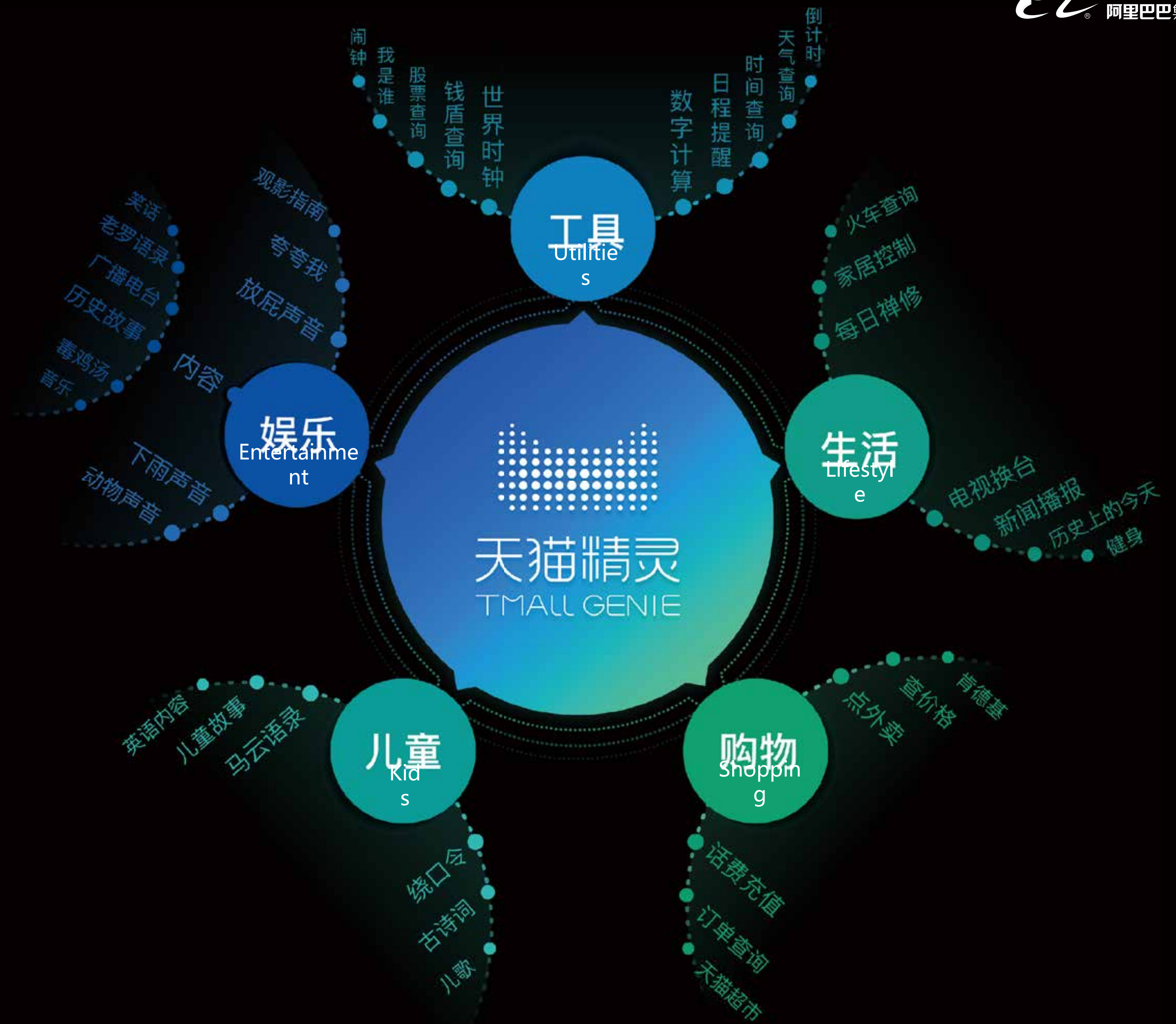
深度学习大幅提升语音识别的准确率，语音逐渐成为一种主要的交互手段



天猫精灵 X1
TMALL GENIE

技能

Skills of TmallGenie



个人语音助手



11:30 午餐

“我要点一份板烧鸡腿堡套餐”、“确认支付”



15:00 机场

“我的航班有延误吗”



8:30 出租车

“我有什么待办事项”、“连接阅江楼语音会议室”



17:00 酒店

“把空调调到26度”、“健身房在哪里？”



8:00 出发

“帮我打辆车去黄龙时代广场”



18:30 饭店

“来一份菲力牛排套餐”、“确认支付”



7:30 早餐

今天有阵雨，出门请带伞。今日要闻：阿里巴巴.....



20:00 健身房

“开始练5分钟腹肌撕裂者”



7:00 起床

已为您打开灯和窗帘，今天您有3项安排，分别是:10点项目会议.....



23:00 晚安

“来点安静的音乐”

自然人机交互的便捷性



1 minute

- 1.Unlock cellphone
- 2.Activate music app
- 3.Search a song
- 4.Choose
- 5.Click play



5 seconds

- 1.“Tmall Genie, I want to listen to ***”

演示结果

一位日本网红杭州旅游
住进了一家神奇的酒店

酒店
Hotel

天猫精灵—阿里巴巴的网红

Tmall Genie - Alibaba's superstar



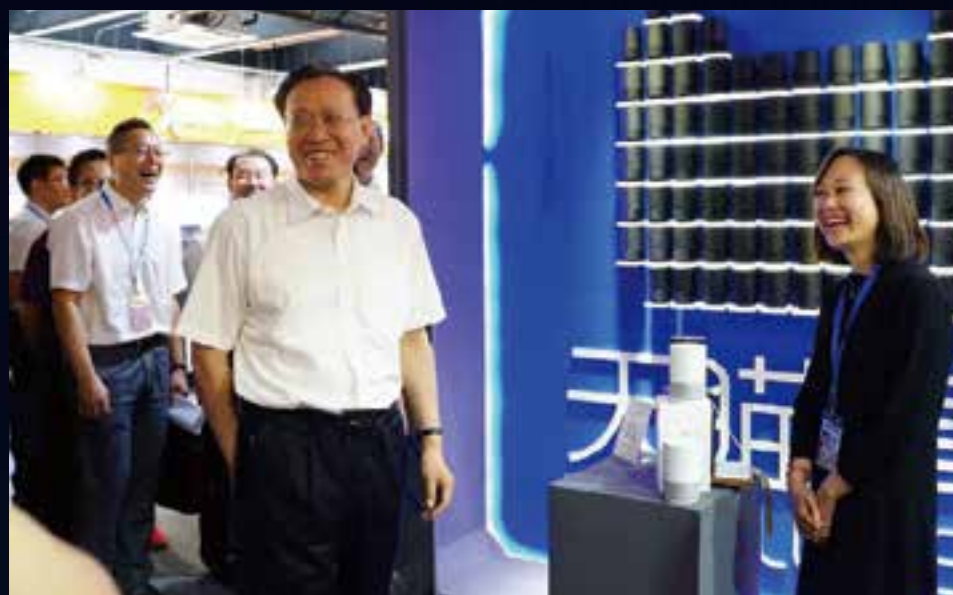
马云带领林郑月娥参观

Jack Ma and Hongkong chief executive Lin Zhengyuee



墨西哥总统参观了解

Jack Ma and President of Mexico



双创周上国家发改委参观

National Development and Reform
Commission of China



科技部秦勇司长亲自体验

Ministry of science and technology Of
China



潘院士称赞天猫精灵

Chinese Academy of Science

销售记录

Sales



双11出货一百万台
奠定行业第一

No.1 in Chinese Market



影音类目销售第一

No.1 in Category

天猫全球狂欢节
2017

截至11月11日8时53分
天猫精灵双11总销量

破100万台

Ali A.I. Labs
阿里人工智能实验室

天猫精灵
TMALL GENIE



我们的技术框架



我们的核心AI能力



声纹识别

全球首创商业化应用



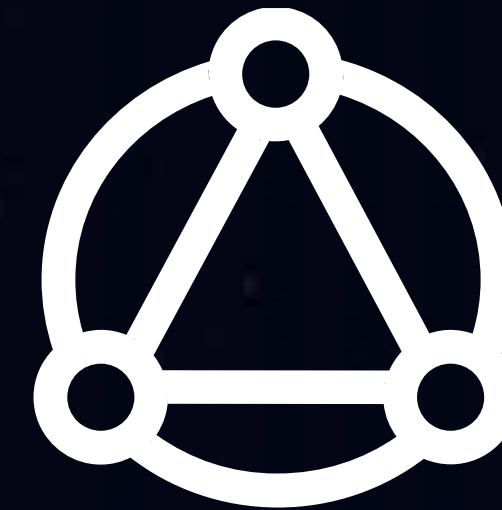
语音识别

远场识别率>97%



自然语言理解

整体理解正确率>97%



知识图谱

上百个领域的知识图谱



云服务

数百个应用服务

旨在打造一套超强AI平台，赋能终端

AliGenie AI开放平台

AliGenie AI Platform

语音识别能力
ASR

语义理解能力
NLP

图像识别能力
Image
Recognition

3D视觉能力
3D Vision

多模态
融合能力
Sensor Fusion

工程化能力
Fast Implementation

硬件终端能力
Hardware Adaptation

产品演进路线



欢迎加入乐观，聪明，有热情的我们

落地AI, 改变世界！

走进阿里妈妈技术

阿里妈妈 靖世

阿里妈妈业务



阿里妈妈是什么？

阿里集团的大数据营销平台

阿里妈妈有哪些产品？

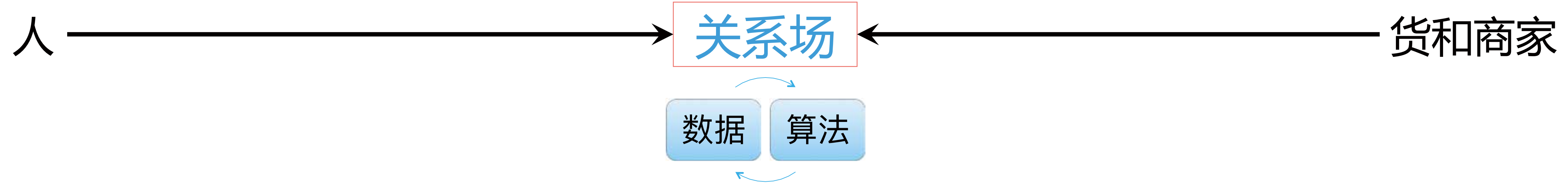
淘宝直通车 智钻 淘宝客 品销宝 达摩盘 UniDesk 等

阿里妈妈的作用

让天下没有难做的营销：服务450万品牌和商家，输送进店流量28万亿，带来47亿笔成交订单
阿里巴巴集团的现金奶牛



通过综合考虑商家、消费者、平台，提供营销价值，创新营销模式



匹配

下一代匹配技术：
任意深度学习+树结构全库检索

模型

互联网模型演化之路：
大规模稀疏非线性MLR
深层用户兴趣分布网络DIN
高级模型服务器AMS和用户浏览
图像感知网络

商业机制和售卖

消费者、平台、商家三赢机制：
Optimized CPC

基础算法：图像&NLP
工程和架构技术：深度学习框架XDL，广告检索引擎
基础平台：风控和反作弊、前端等

大规模稀疏非线性模型 MLR

CTR预估经典方法：大规模特征+LR

- 大规模：样本量大，特征维度高，稀疏特征
- LR：广义线性模型，大量人工特征工程，L1正则做特征选择

问题：

- 人工能力有限，很难对非线性模式挖掘完全充分
- 依赖人力和领域经验，方法推广到其它问题的代价大：不够智能
- 已有的非线性模型不适用：Kernel方法，Tree-based方法（GBDT），FM类方法

我们提出分片线性学习算法MLR（2011尝试，2012上线）：

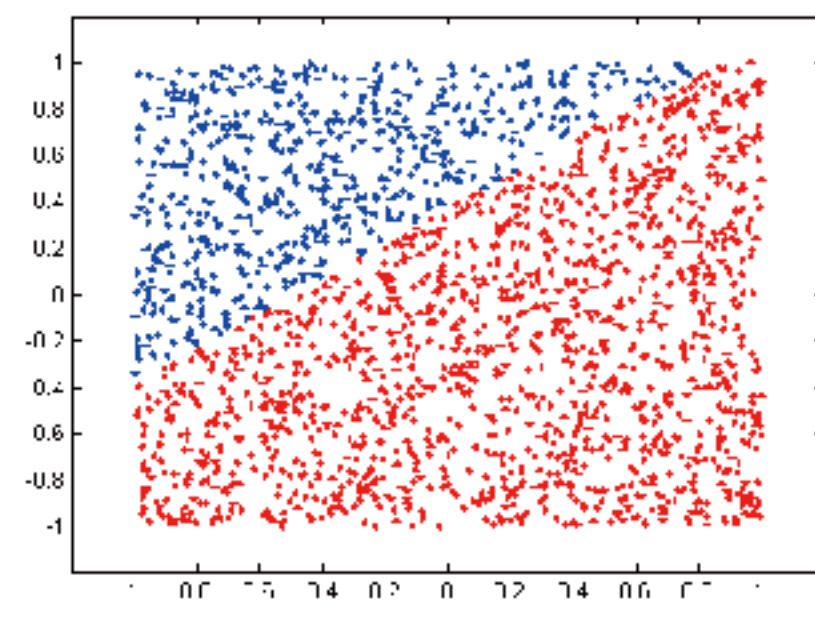
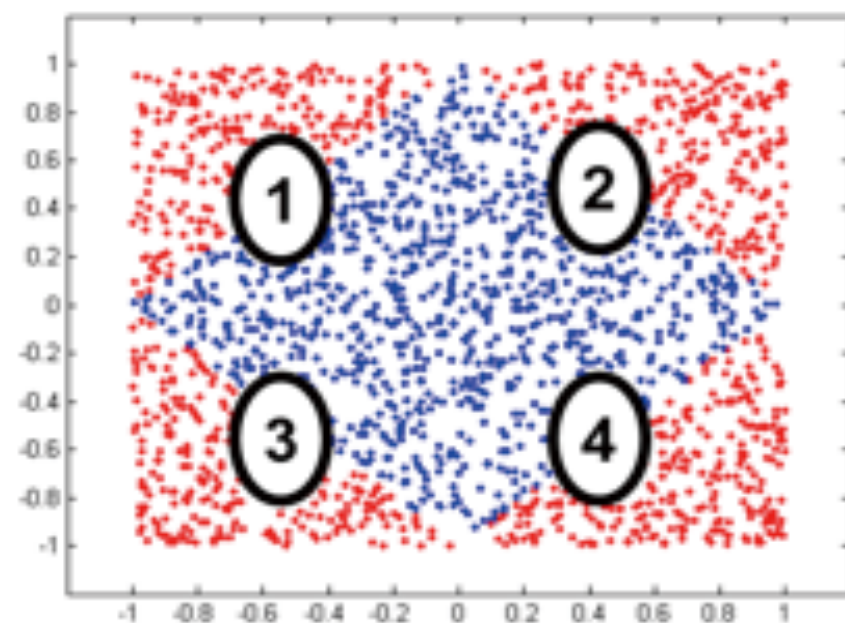
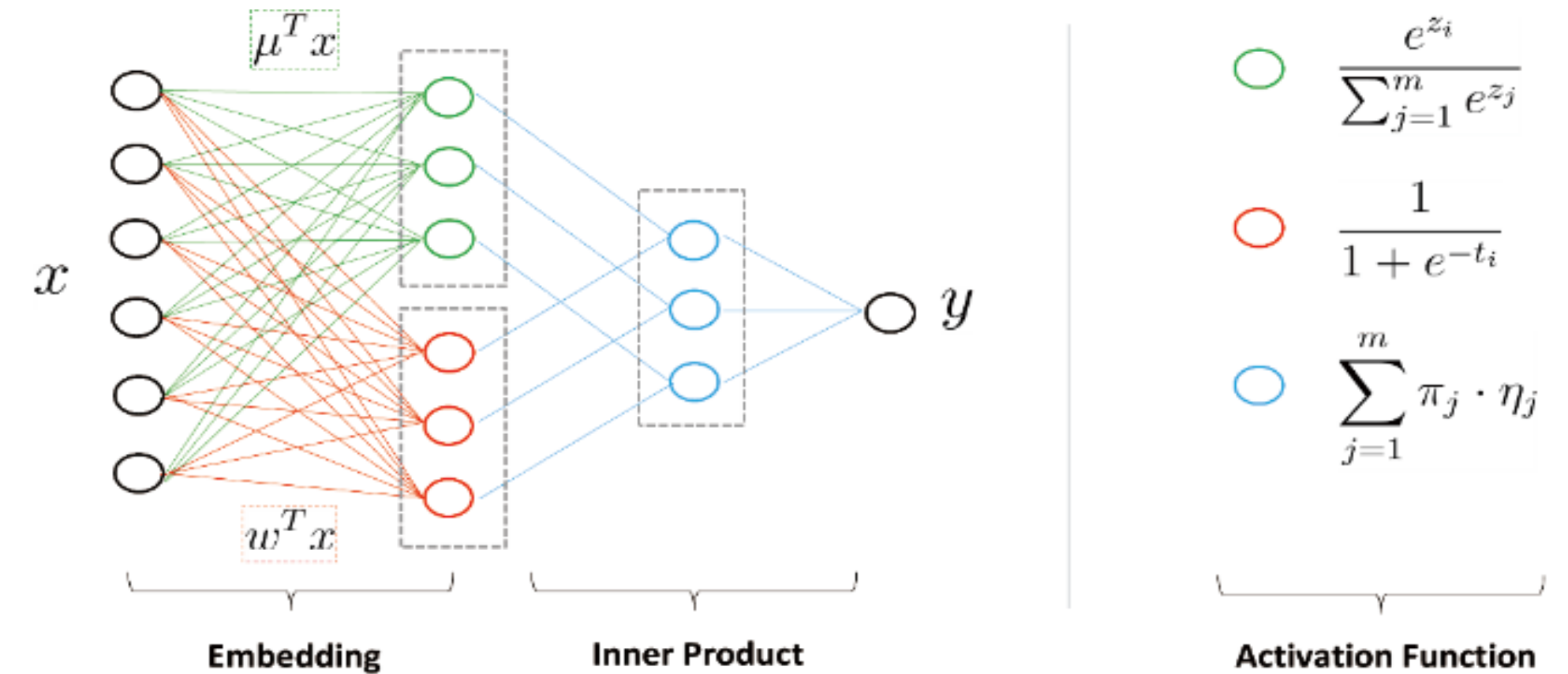
- 任意强非线性拟合能力；
- 分片控制模型复杂度，平衡过拟合欠拟合，控制每分片足够多样本增强泛化能力；
- 特征选择能力，分组稀疏L21正则

分而治之：

业务主要用的形式：

$$f(x) = g\left(\sum_i \pi_i(x, \mu) \eta_i(x, w)\right)$$

$$f(x) = \sum_{i=1}^m \frac{e^{\mu_i \cdot x}}{\sum_{j=1}^m e^{\mu_j \cdot x}} \cdot \frac{1}{1 + e^{-w_i \cdot x}} \quad f(x) = \left(\sum_{i=1}^m \frac{e^{\mu_i \cdot x_u}}{\sum_{j=1}^m e^{\mu_j \cdot x_u}} \cdot \frac{1}{1 + e^{-w_i \cdot x_a}} \right) \cdot \frac{1}{1 + e^{-w \cdot x_2}}$$



深层用户兴趣分布网络 Deep Interest Network

深度学习通用做法：

- Embedding嵌入技术
- 多个行为聚合成一个用户向量

问题：

- 用户兴趣多样，而一个向量表示能力受限

我们提出DIN (2016-2017)：

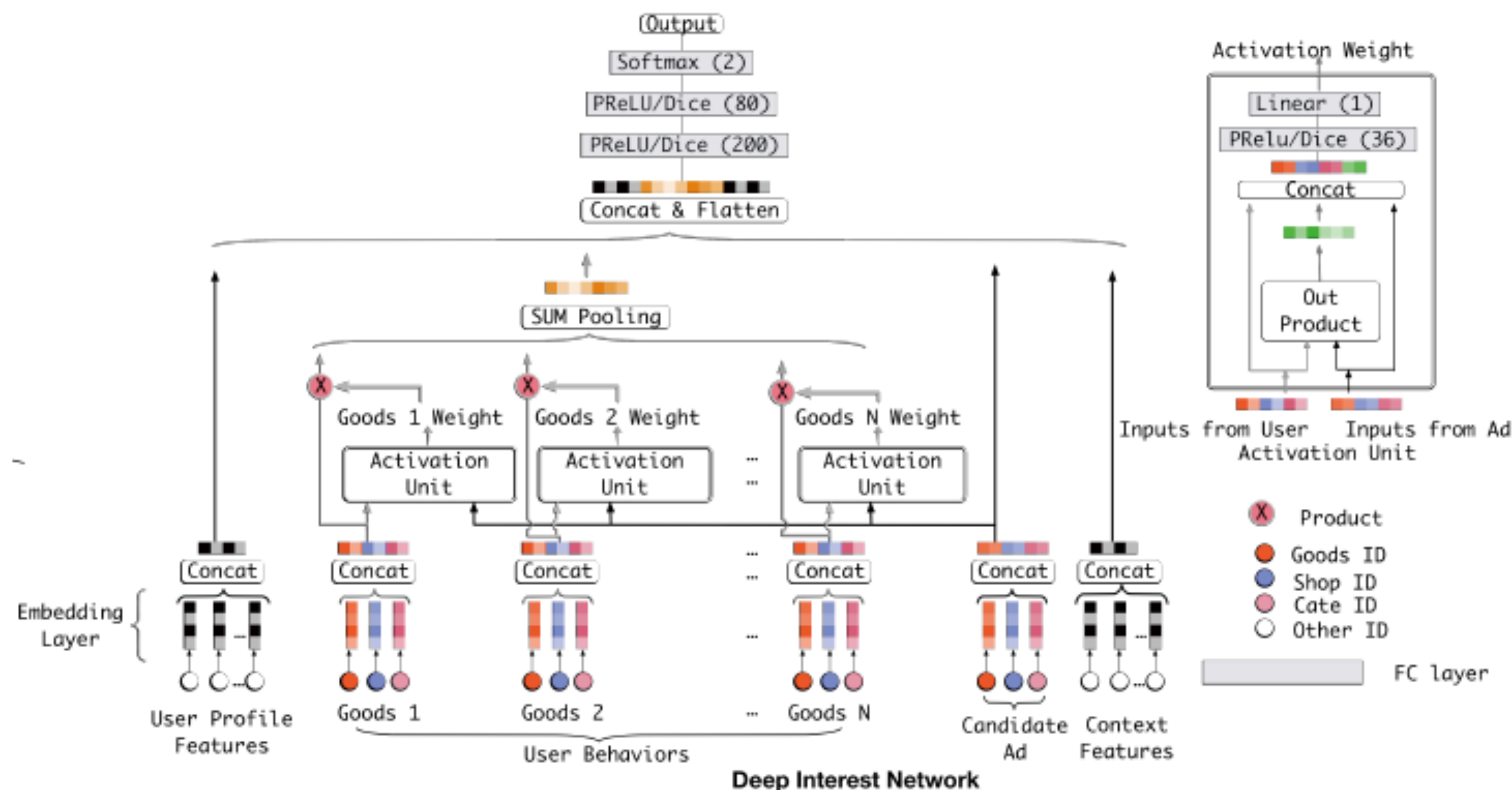
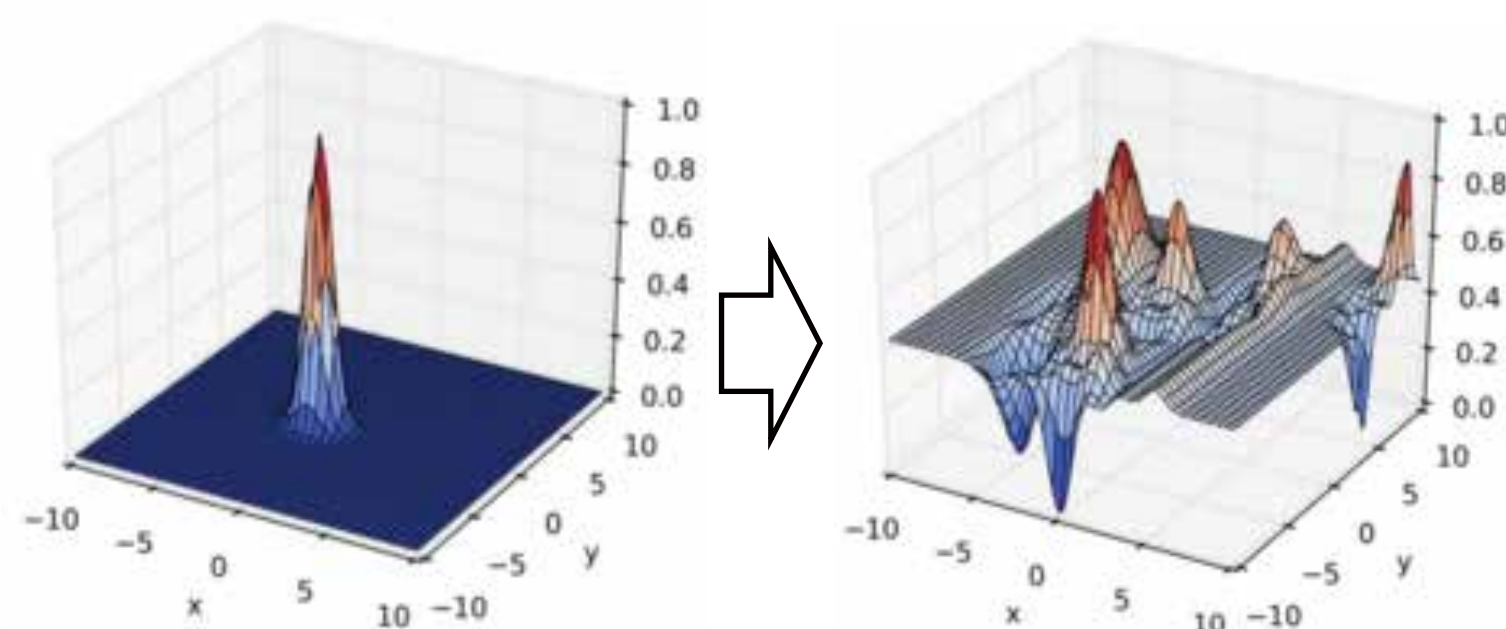
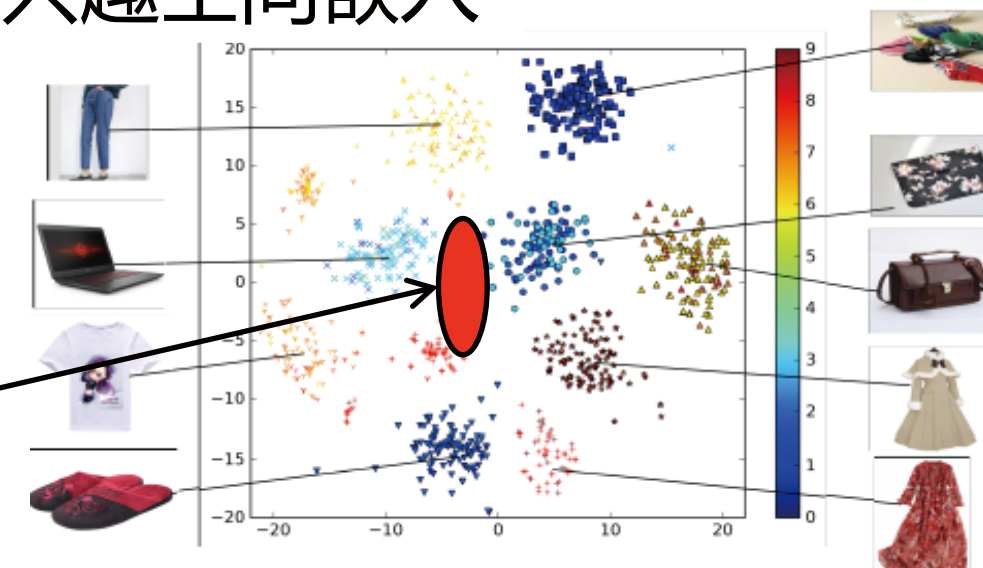
- 我们提出用户兴趣应该是多峰分布，而非一个向量点
- 当用户浏览某个商品时，只有其中部分（通常一个）兴趣被激发

设计：根据要预估的商品去反向挑选行为子序列去组成用户向量
(类似Attention机制)

用户兴趣分布变成非参数多峰形式，表示灵活性大大增加

兴趣空间嵌入

用户兴趣点
内积计算



超越参数分布PS范式，模型分布范式初探和联合训练服务AMS

跨媒介信息：

用户浏览商品id → 浏览图片

困难点：

一个id 变成 一个图片

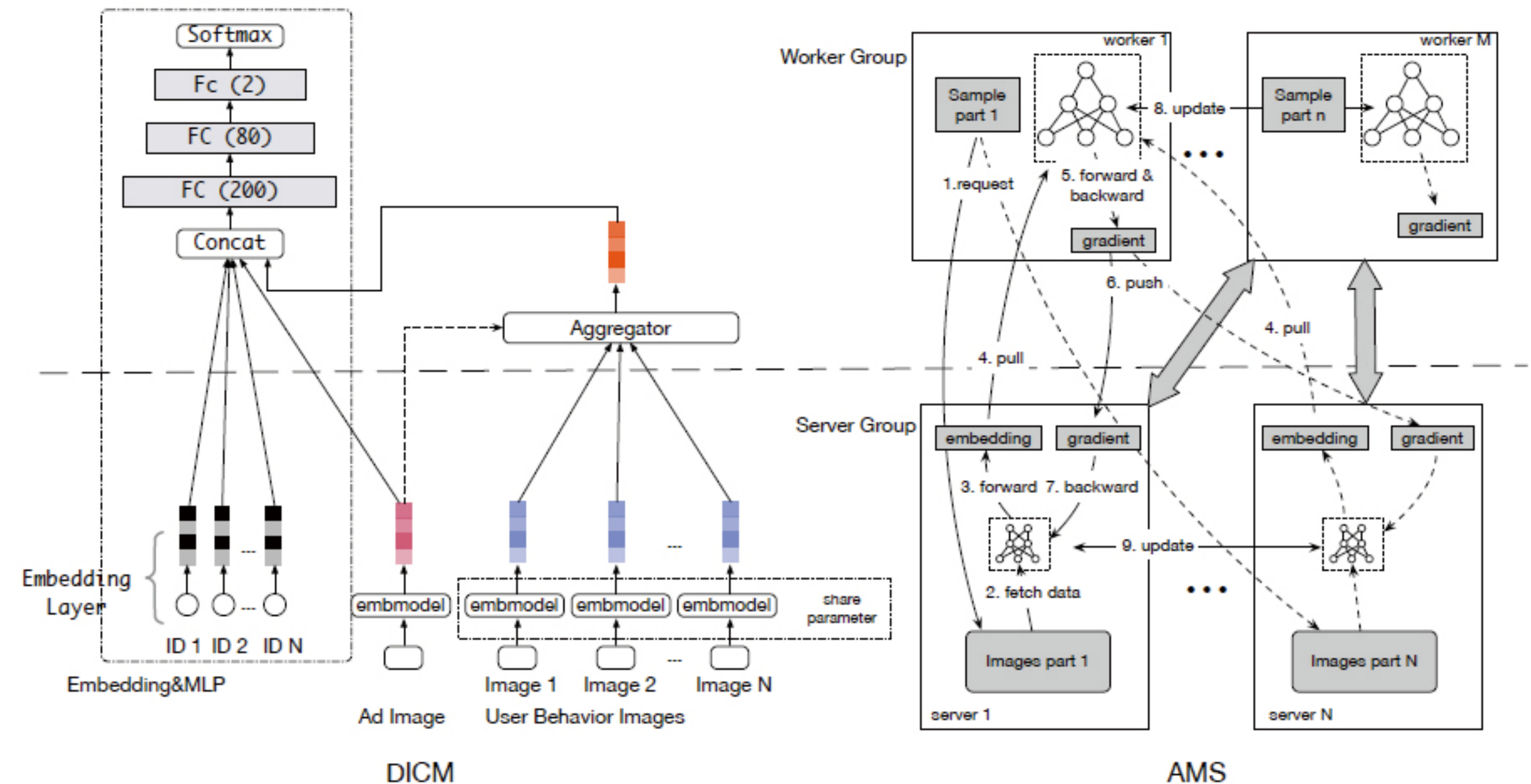
百亿样本，一个样本数十到数百张图片

突破点： Unique图片远没有这么多，可否节省计算资源？

超越参数分布范式：

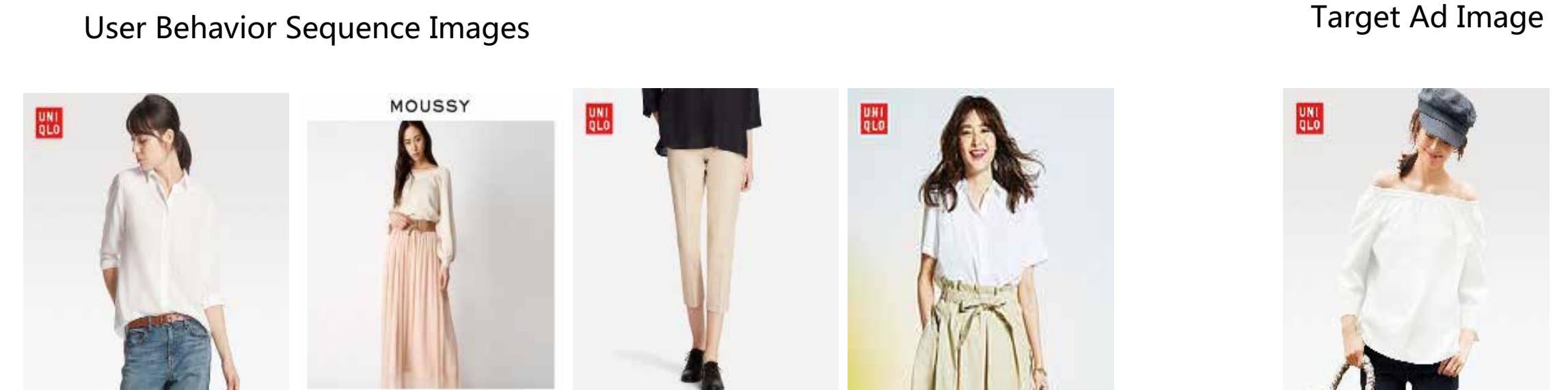
模型分布范式+联合训练服务器Advanced Model Server (2017)

使得这个任务变得可能



Powered by :

1. XDL: 阿里妈妈建立的一套完整的超大规模深度学习框架
2. Alimama Vision (OCR算法2016年6月刷新ICDAR数据集世界最好成绩)



探索下一代匹配&推荐技术

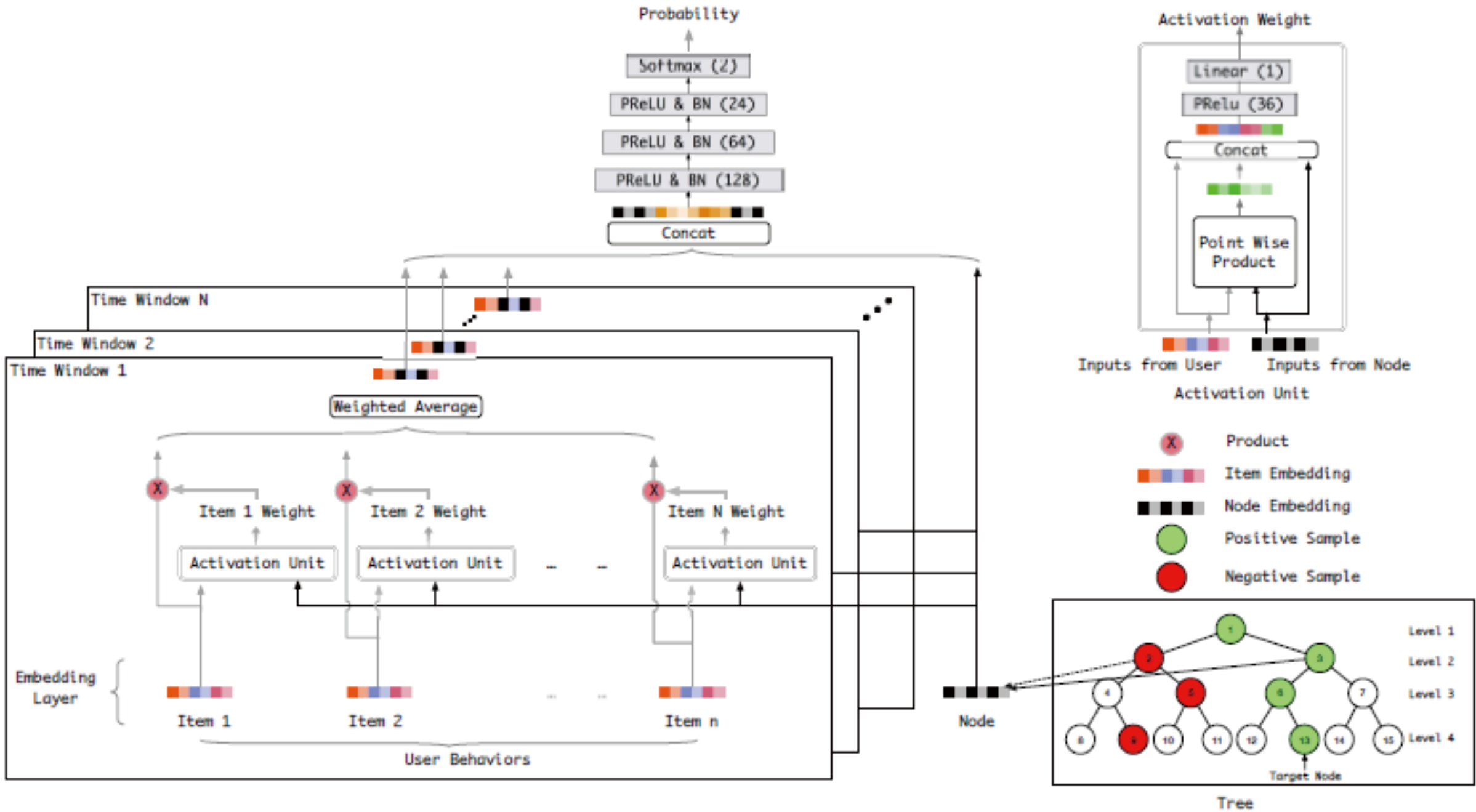


- 第一代：启发式统计规则式，例如Item-CF
- 第二代：内积模型+向量检索引擎 ← 高级深度学习无法应用
- 第三代：任意高级深度学习+树结构全库检索引擎

树结构全库检索
深度网络

对全库构建由粗到细的类别树：例如10亿物料需要30层二叉树
从粗概念判断，逐步细化，直到挑选到具体商品
挑选最优：判断10亿次 → 30次

	Recall	新类目Recall
Item-CF	6.95%	1.06%
Youtube 内积DNN	7.58%	3.09%
Ours	12.37%	4.82%



Optimized CPC

背景：综合提升 (a) 广告主收益，(b) 用户端gmv，(c) 商业端rpm

思路: 帮助广告主提高优质流量出价，也同时能增加rpm和gmv [1]

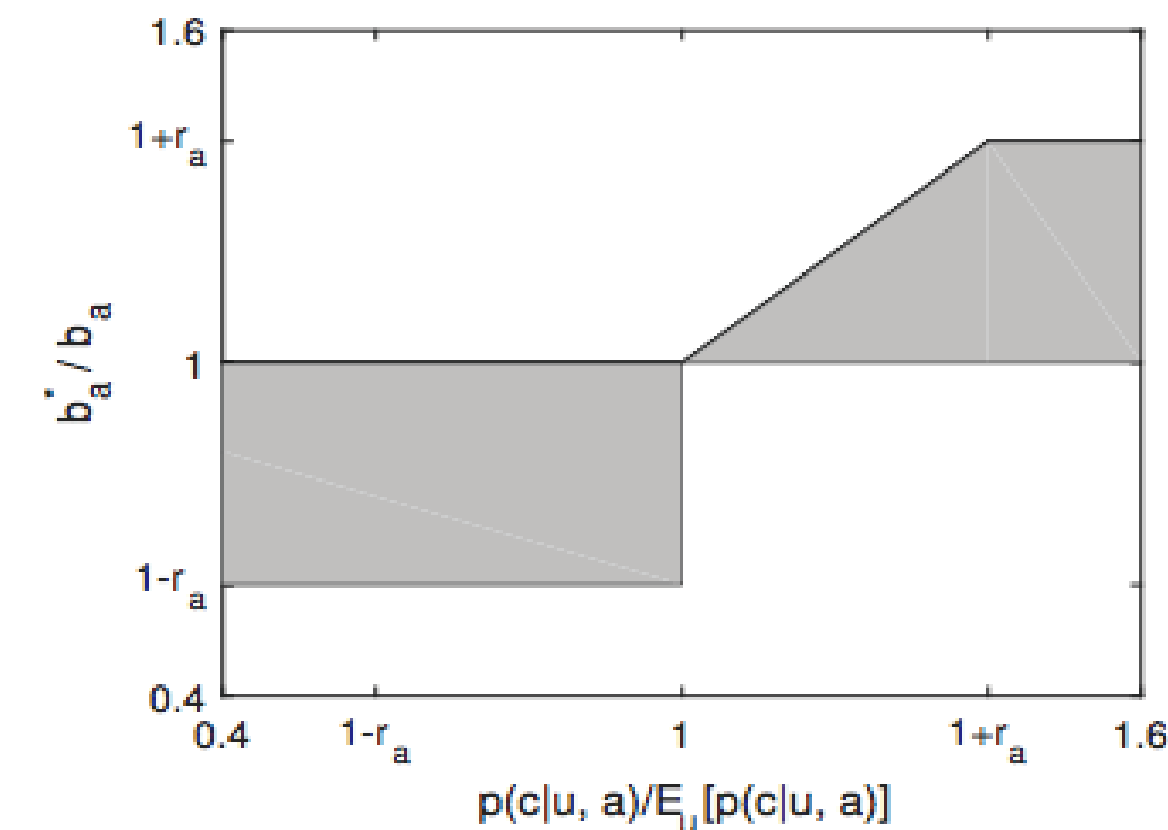
- 广告主授权优化目标和调价范围，确定调价上下限；
- 每组具体价格，ecpm排序展现后对应不同的平台指标
- 最优化所要的平台指标，求解具体调价：

$$\begin{aligned} \max_{b_1^*, \dots, b_n^*} \quad & f(k, b_k^*) \\ \text{s.t.} \quad & k = \operatorname{argmax}_i \text{pctr}_i * b_i^* \\ & l(b_i^*) \leq b_i^* \leq u(b_i^*), \forall i \in A \end{aligned}$$

$$f_1(k, b_k^*) = \text{pctr}_k * \text{pcvr}_k * v_k,$$

$$f_2(k, b_k^*) = \text{pctr}_k * \text{pcvr}_k * v_k + \alpha * \text{pctr}_k * b_k^*$$

- 下一代广告主和平台优化技术：机制和强化学习 [2,3]



[1] Han Zhu, Junqi Jin, Chang Tan, Fei Pan, Yifan Zeng, Han Li, Kun Gai. Optimized Cost per Click in Taobao Display Advertising. In KDD 2017.

[2] Junqi Jin, Chengru Song, Han Li, Kun Gai, Jun Wang, Weinan Zhang. Real-Time Bidding with Multi-Agent Reinforcement Learning in Display Advertising. <https://arxiv.org/abs/1802.09756>

[3] Di Wu, Xiujun Chen, Xun Yang, Hao Wang, Qing Tan, Xiaoxun Zhang, Kun Gai. Budget Constrained Bidding by Model-free Reinforcement Learning in Display Advertising. <https://arxiv.org/abs/1802.08365>

这是一个什么样的团队



我们相信什么：

通过技术前进带来业务红利的同时，在人类科技树上做点不一样的东西。

团队愿景：

打造世界超一流的技术团队，超一流的业务团队。

让你变成更好的自己。

如果你对技术有热情，想做点不一样的东西引领前沿，想看到不一样并且更好的自己，欢迎加入我们！

从人工智能到机器智能

达摩院机器智能实验室 智捷

iDST → MIT

- iDST
 - Institute of Data Science and Technologies
- Machine Intelligence Technologies
 - 机器学习智能技术研究院

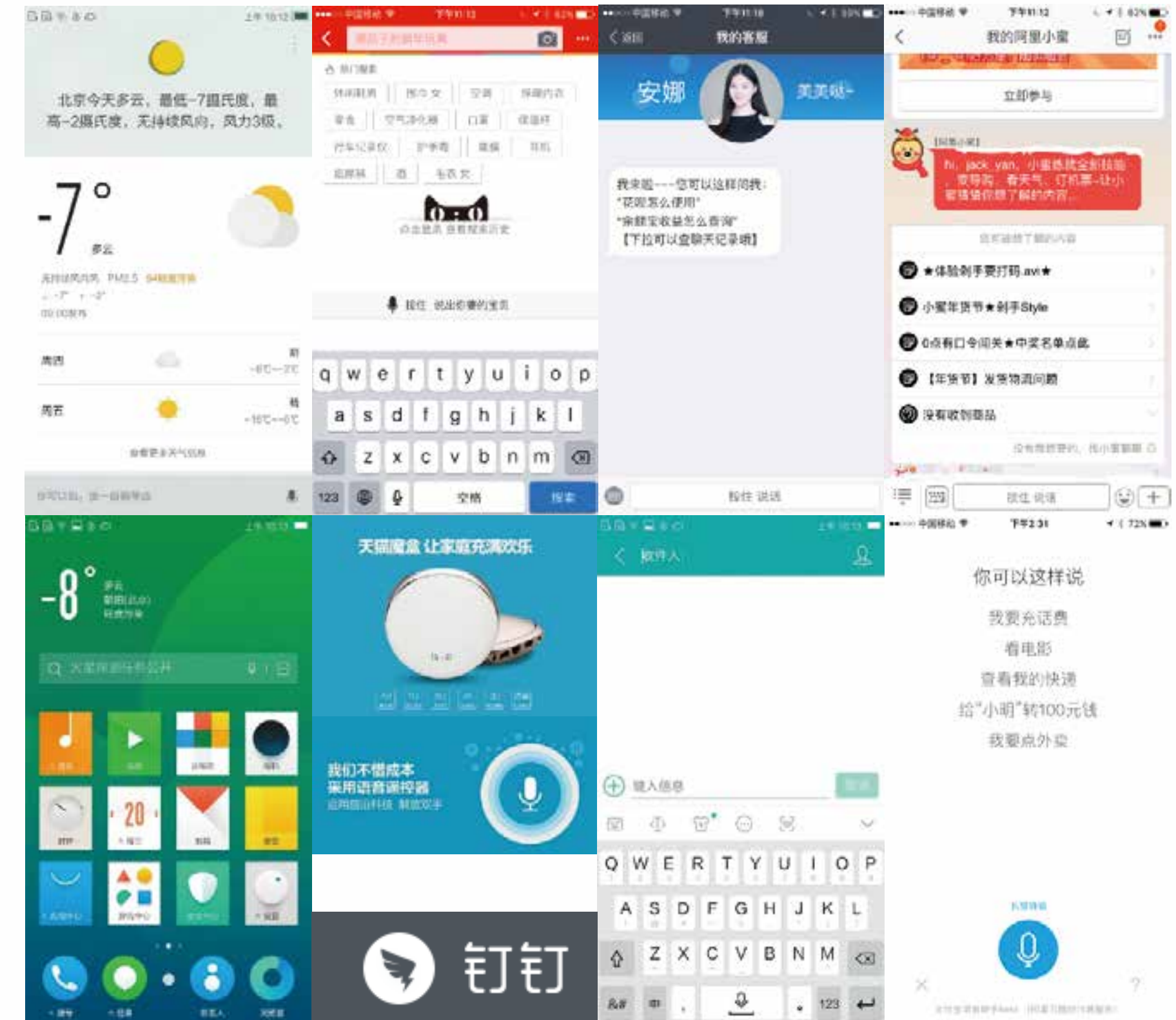
AI = Sensing + Learning + Optimization

- Sensing technologies
 - Speech/voice technologies
 - Image/video technologies
 - Natural language processing (NLP)
- Machine learning
 - Deep learning
 - Reinforcement learning
- Optimization & Decision making
 - Predictive inventory optimization
 - Delivery assignment optimization
 - Manufacturing scheduling
 - Predictive maintenance

Speech Technologies

Our Works – Support Internal Customers

- Core speech technology team at Alibaba
- Alibaba call centers
 - 10k+ phone calls, 100k+ hours of speech per day
 - AliMe
- Alibaba super-apps & devices
 - Taobao, Alipay, Dingding, ...
 - AliOS phones / Tmall TV / Tmall Genie
 - Youku video inspection (via speech)



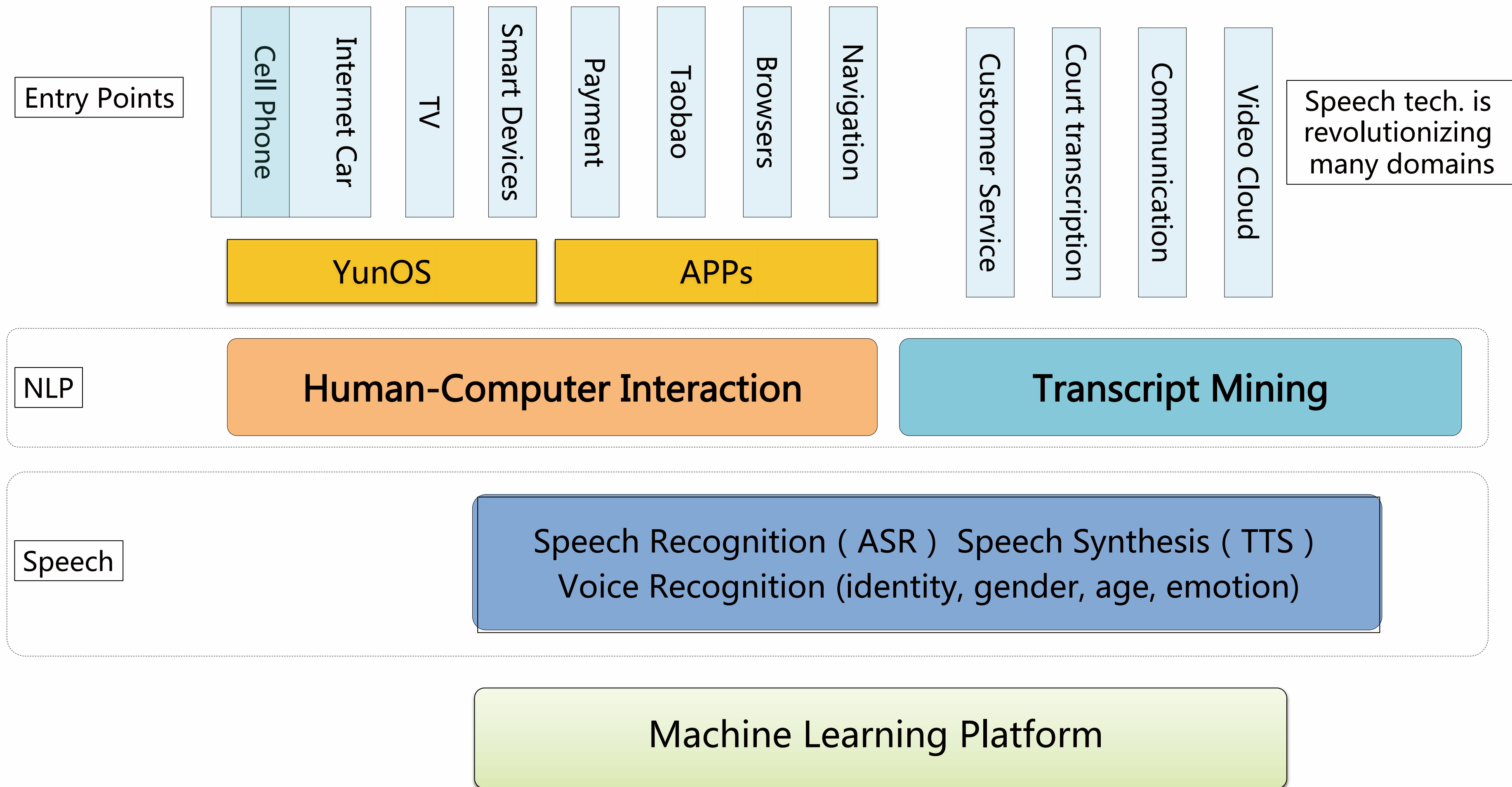
Our Works – Serve External Customers



Speech Interactive Intelligence



Speech based HCI



Core Technologies

- Speech Recognition
 - Big data with good coverage of accents/dialects
 - Deep learning
 - Speech adaption technologies



- Speech Synthesis
 - Mature technologies, but difficult to make it perfect

Core Technologies

- Natural language understanding (NLU)
- Dialogue

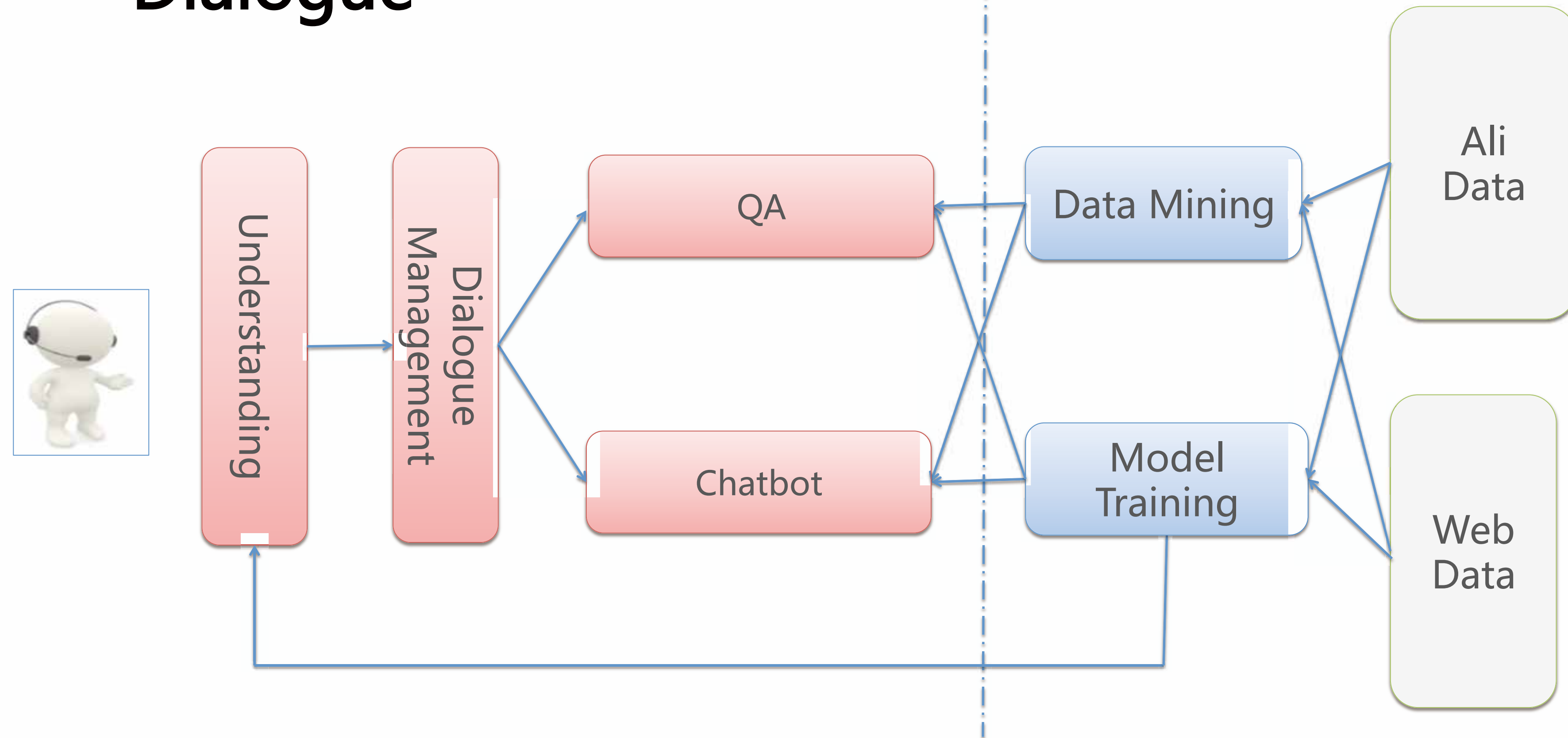
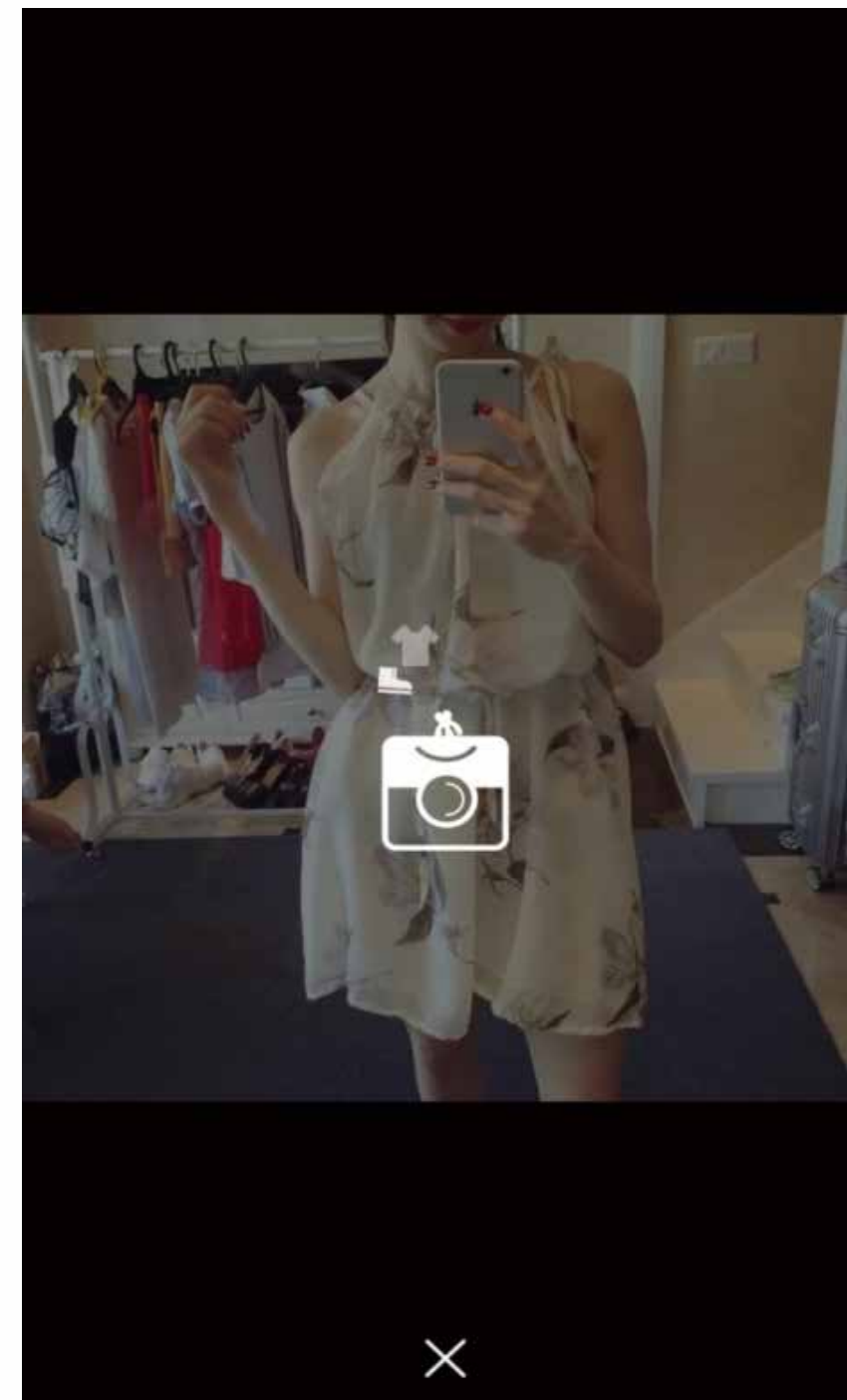


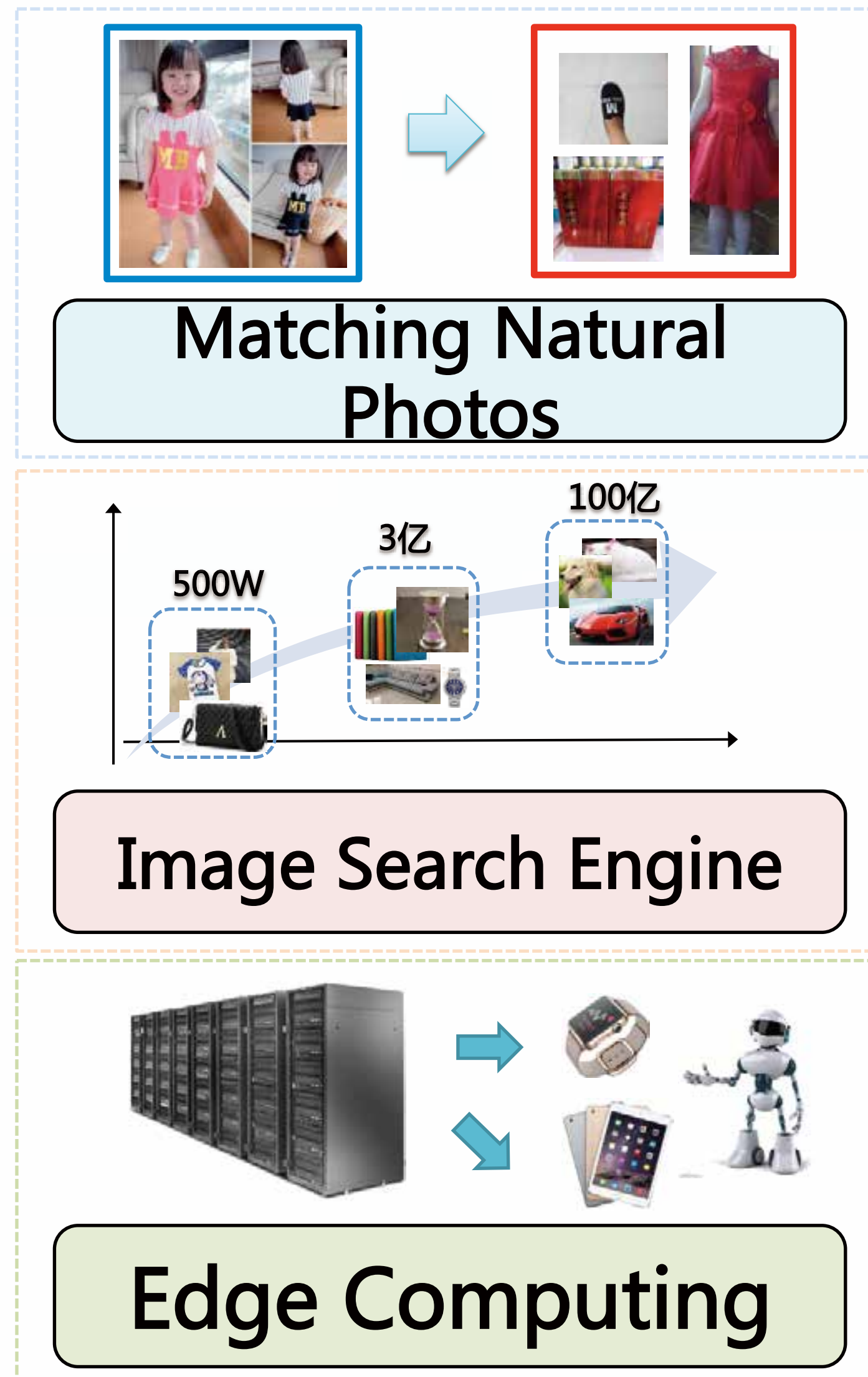
Image Technologies

Image Based Search (拍立淘)

- Based on deep learning technologies
- Index 3 billion images, latency = 0.2s
- 10 million daily active users



Core Technologies



- **Acura**
 - Detection + DL features
 - Personalization
- **Billionaire**
 - Data compression
 - Better indexing scheme
- **MobileDL**
 - Model compression

Cross Media Retrieval

- Search images by texts

dress

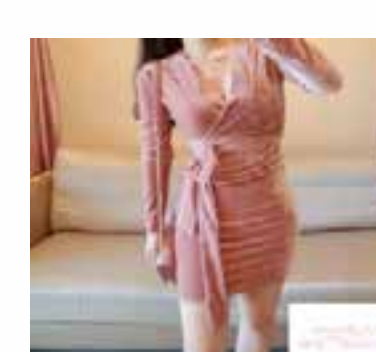


Red dress

Cross
Media
Retr.



V collar short
dress



Red V collar
short dress



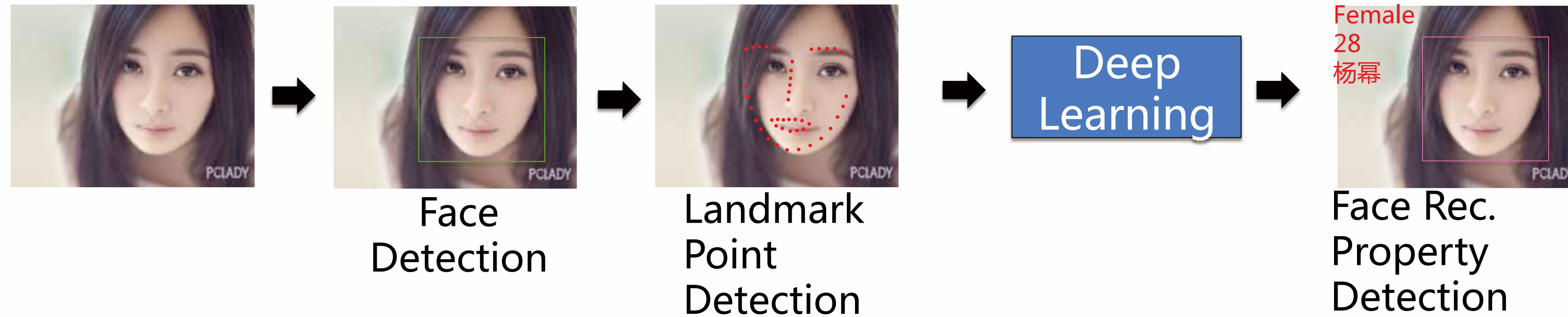
OCR



Applications :

- Image content extraction
- Filtering out advertisements

Face



Applications

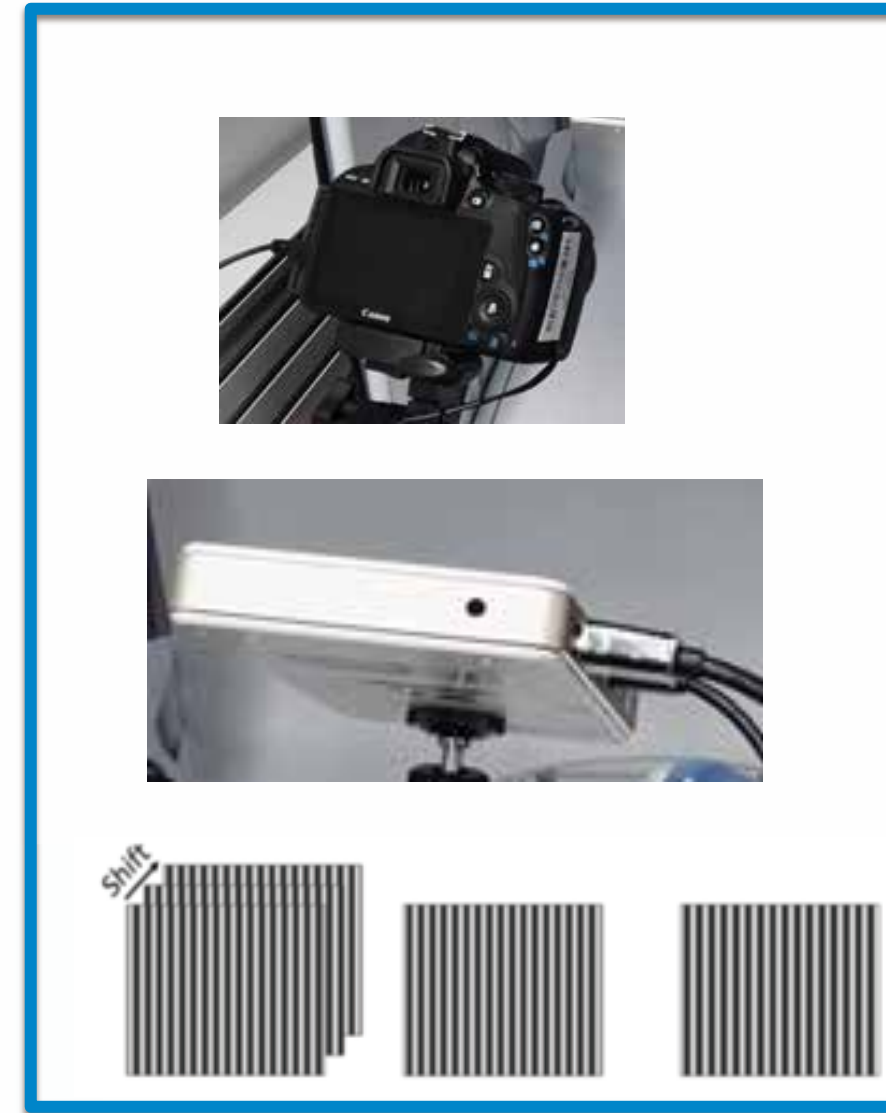
- Identity verification
- Video meta data extraction

3D Modeling

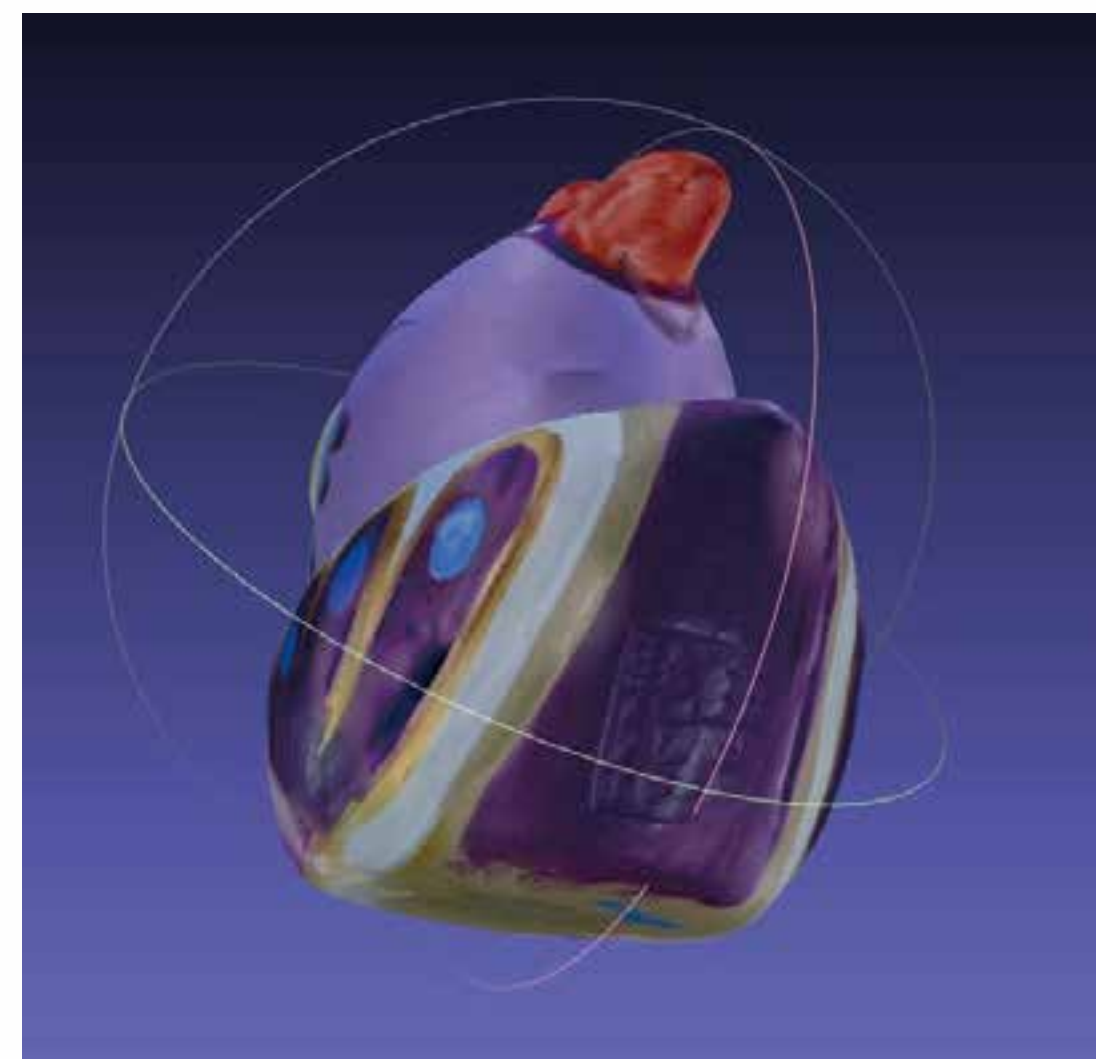
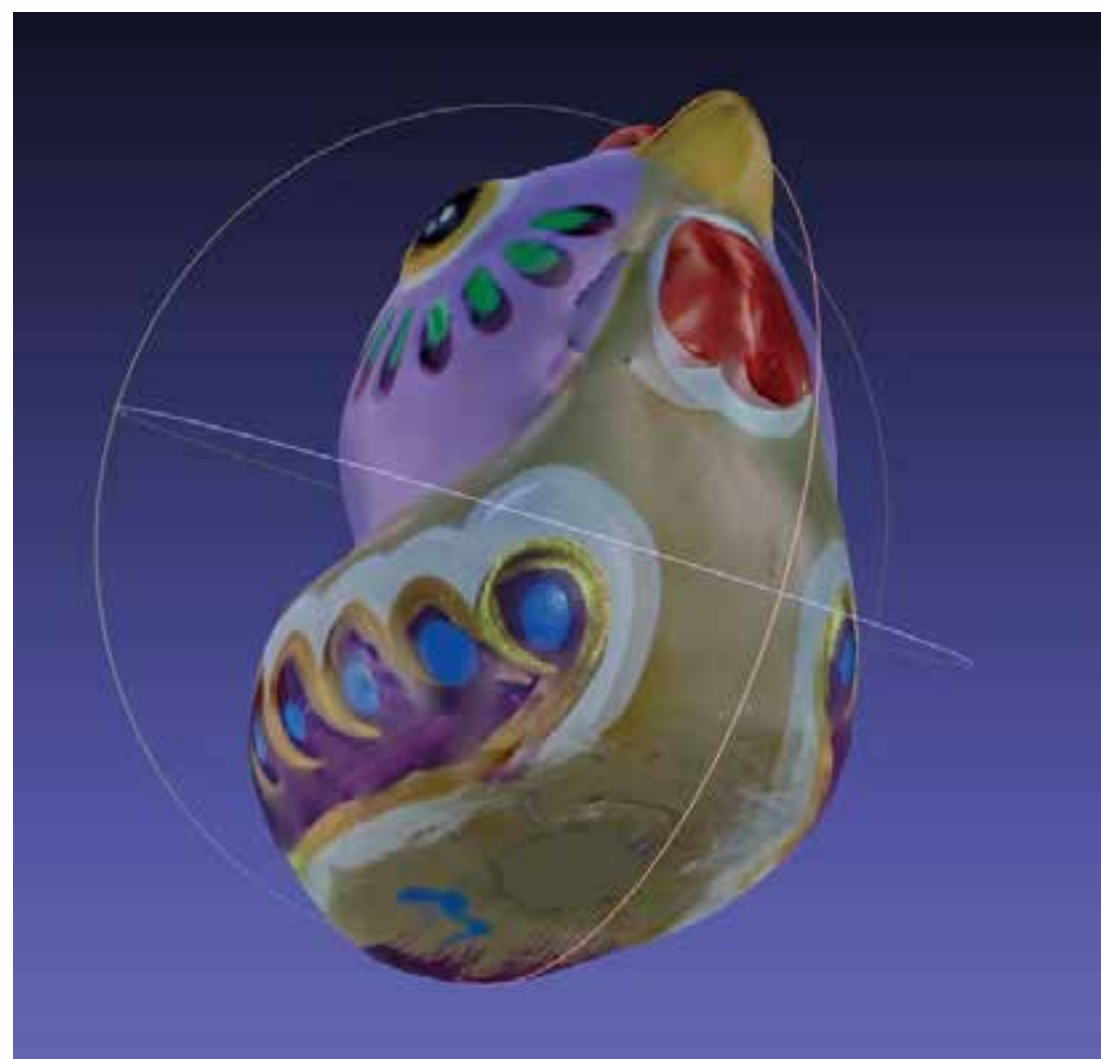
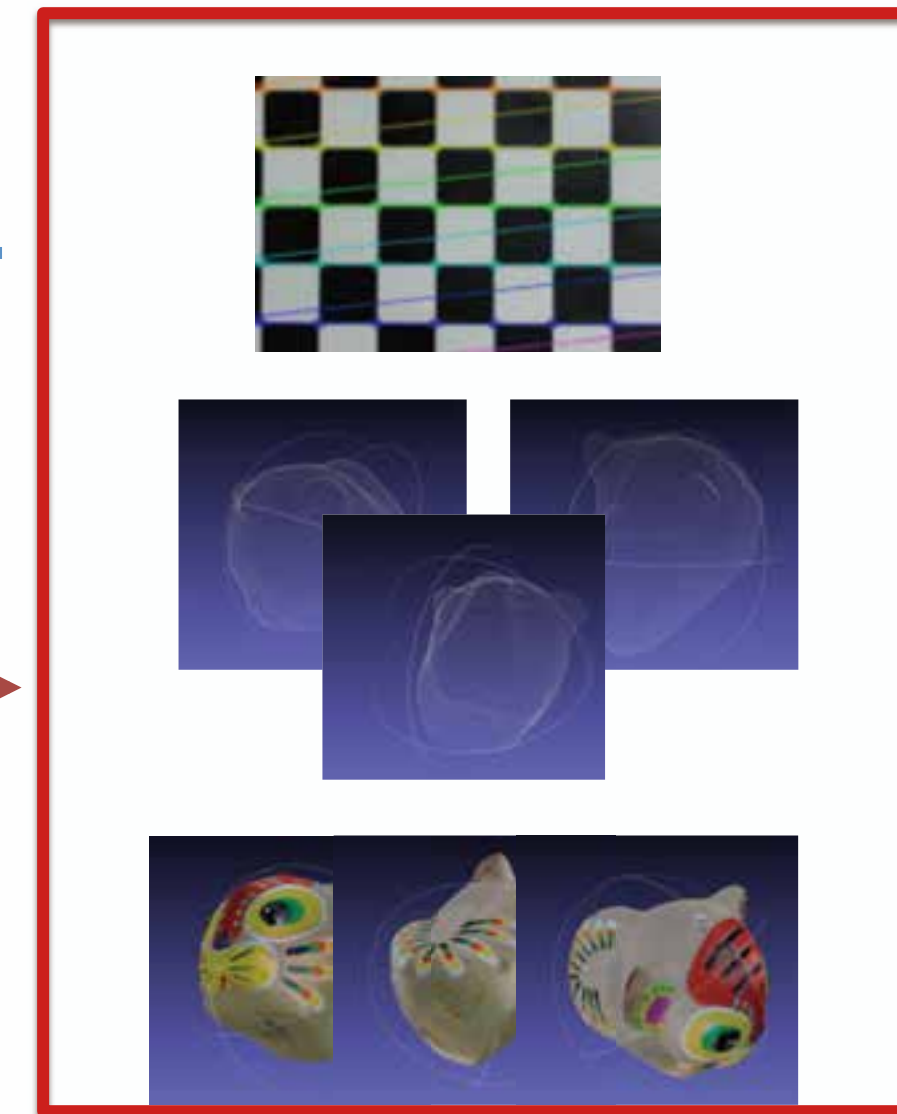
Application



Low Cost Device

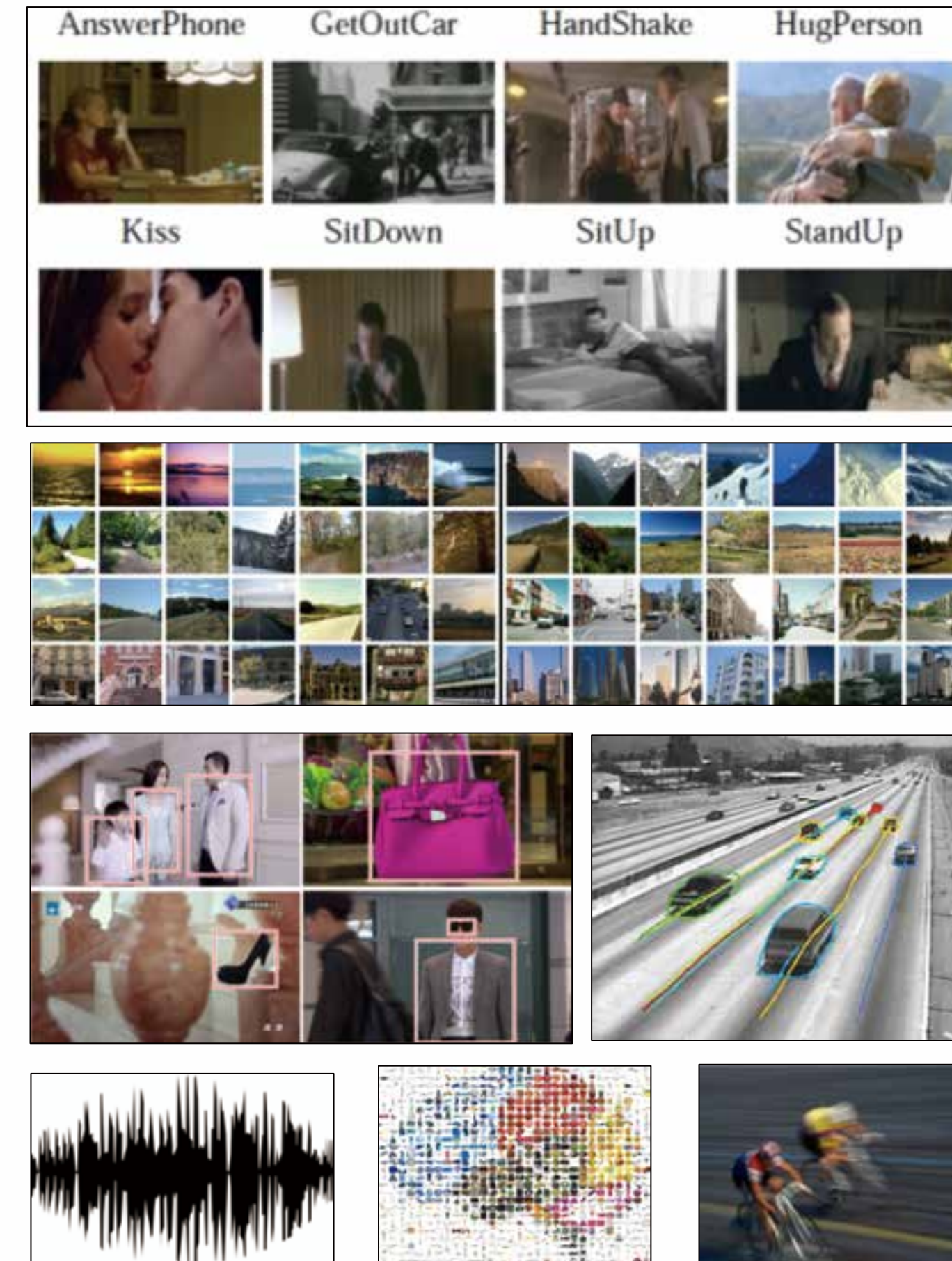
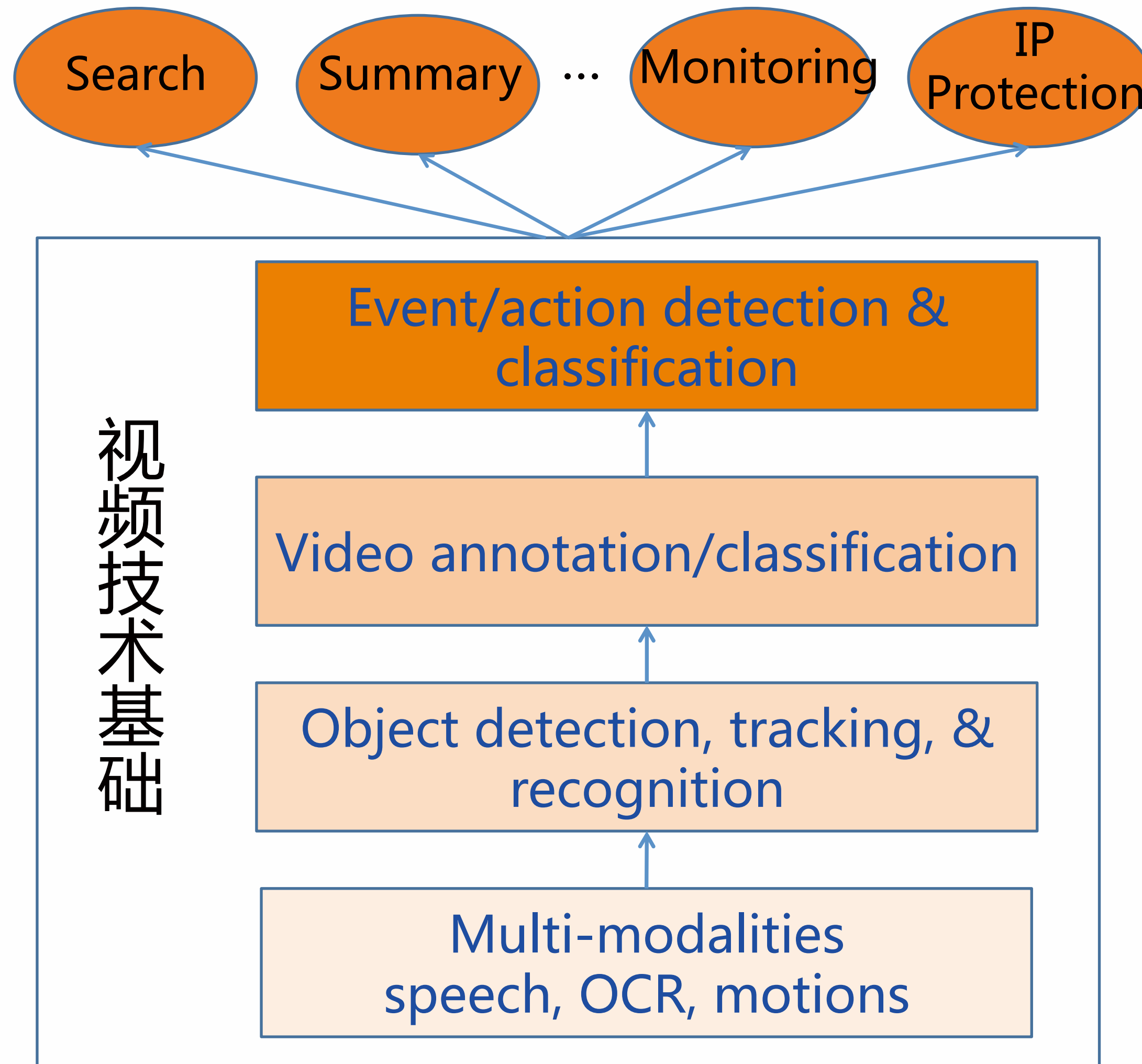


High Resolution Modeling



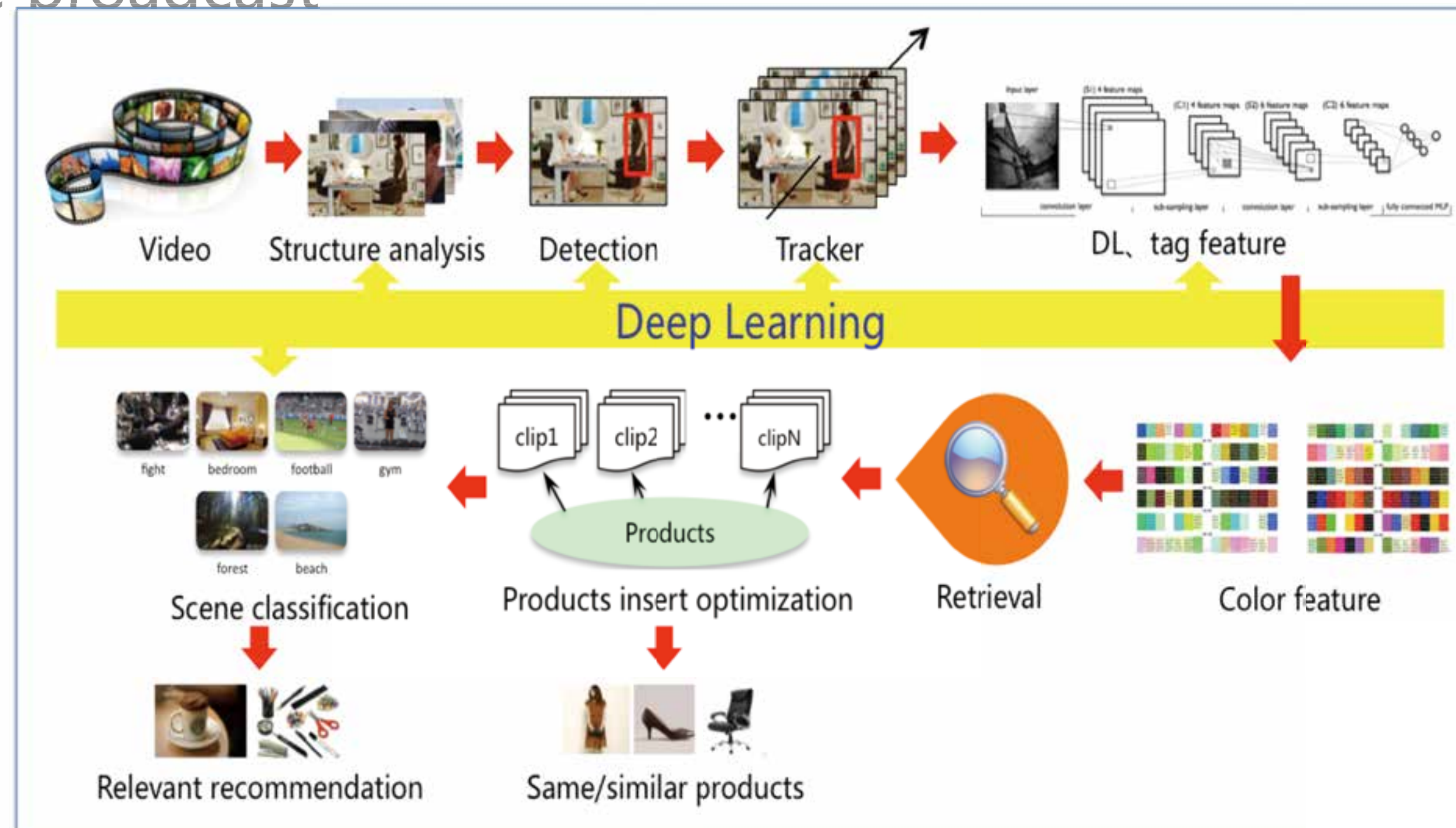
Video Technologies

Core Video Technologies



Venus: Video Analysis System

- Integrate detection, tracking, recognition algorithms
- Applications:
 - video content understanding
 - live broadcast



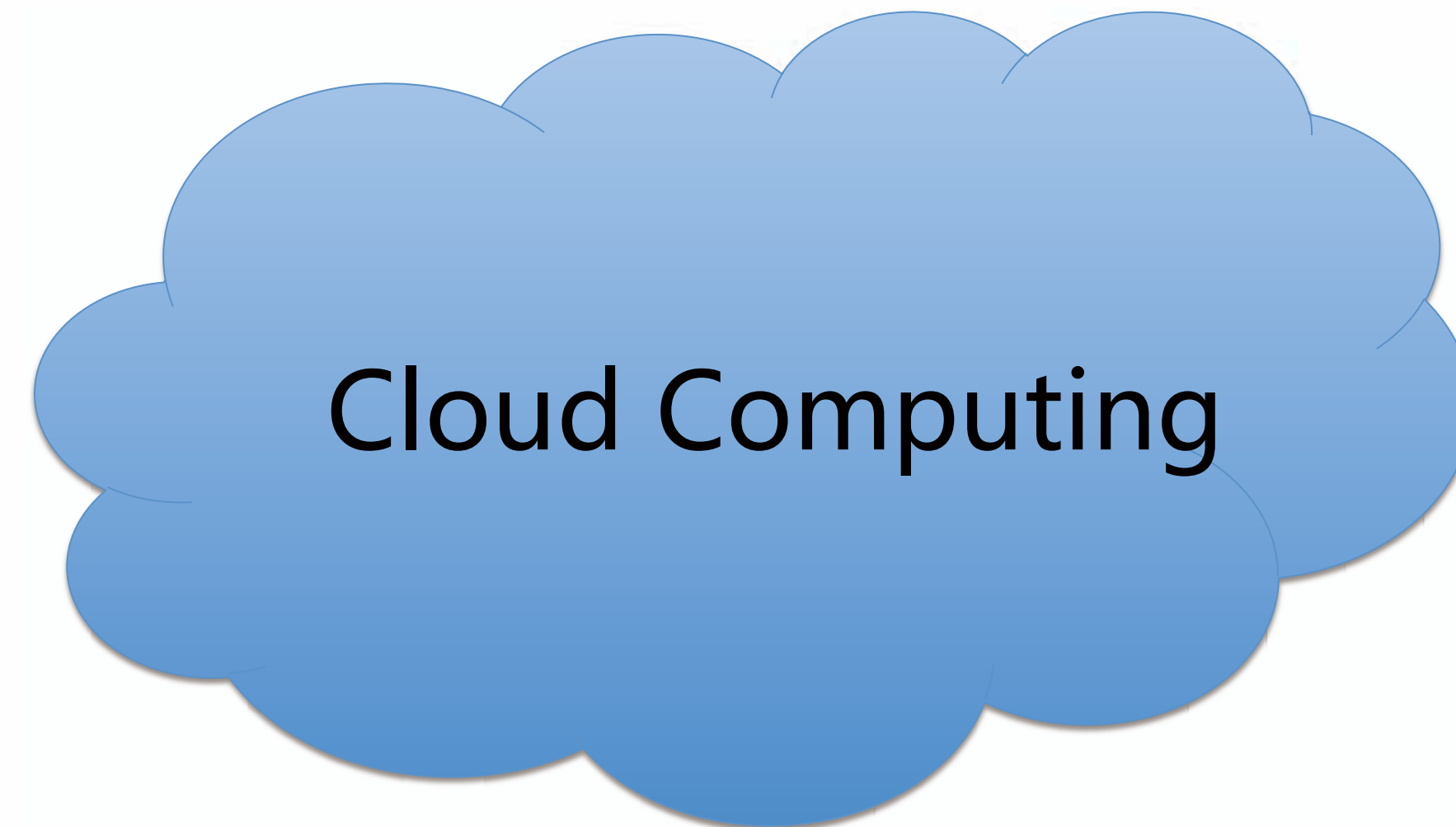
High Performance Computing



GPU



FPGA



SoC + IP



Edge computing: ARM+GPU、FPGA、SoC



Scanner



Camera



AR



ADAS



VR



Drone

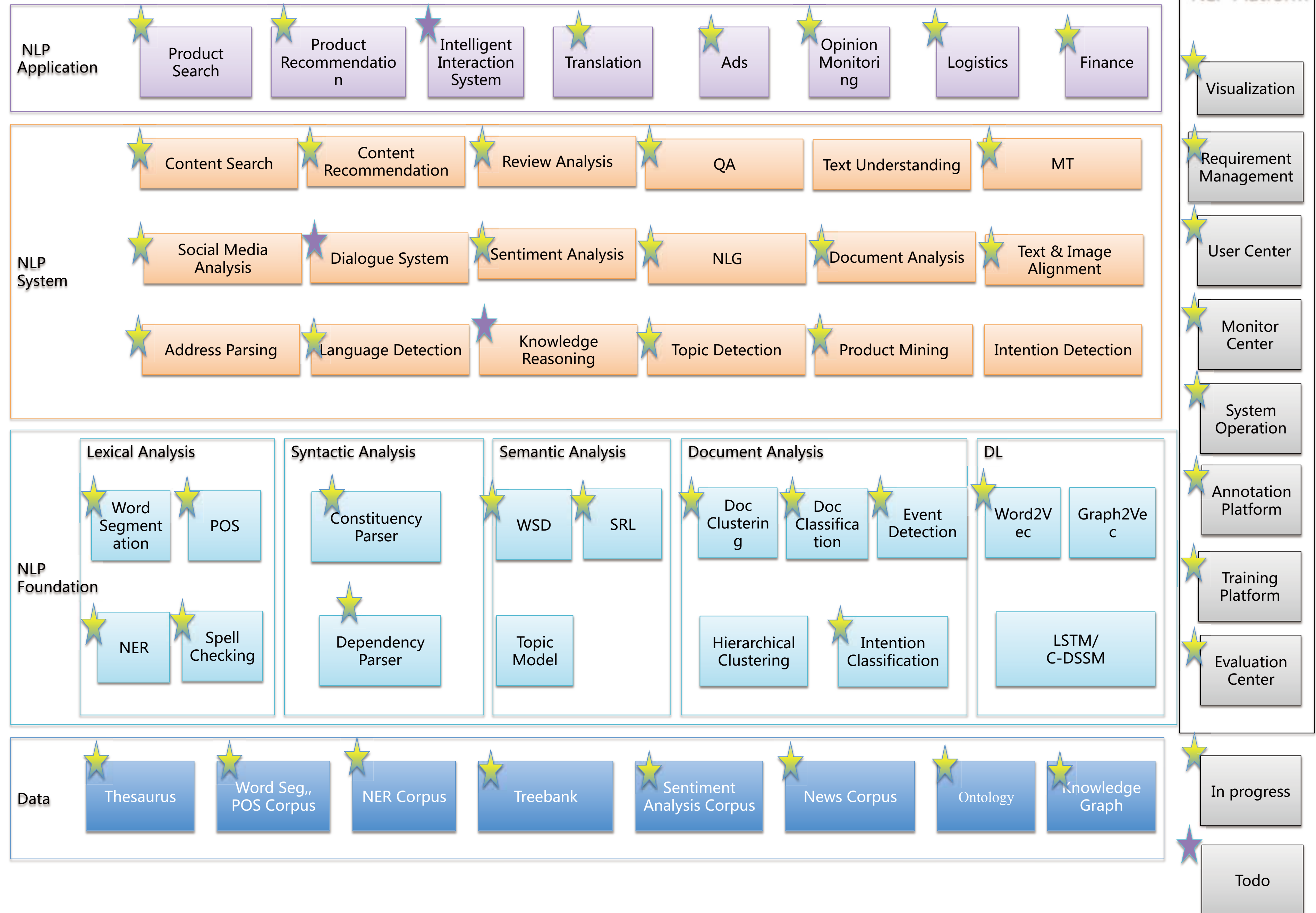


Robot

Natural Language Processing

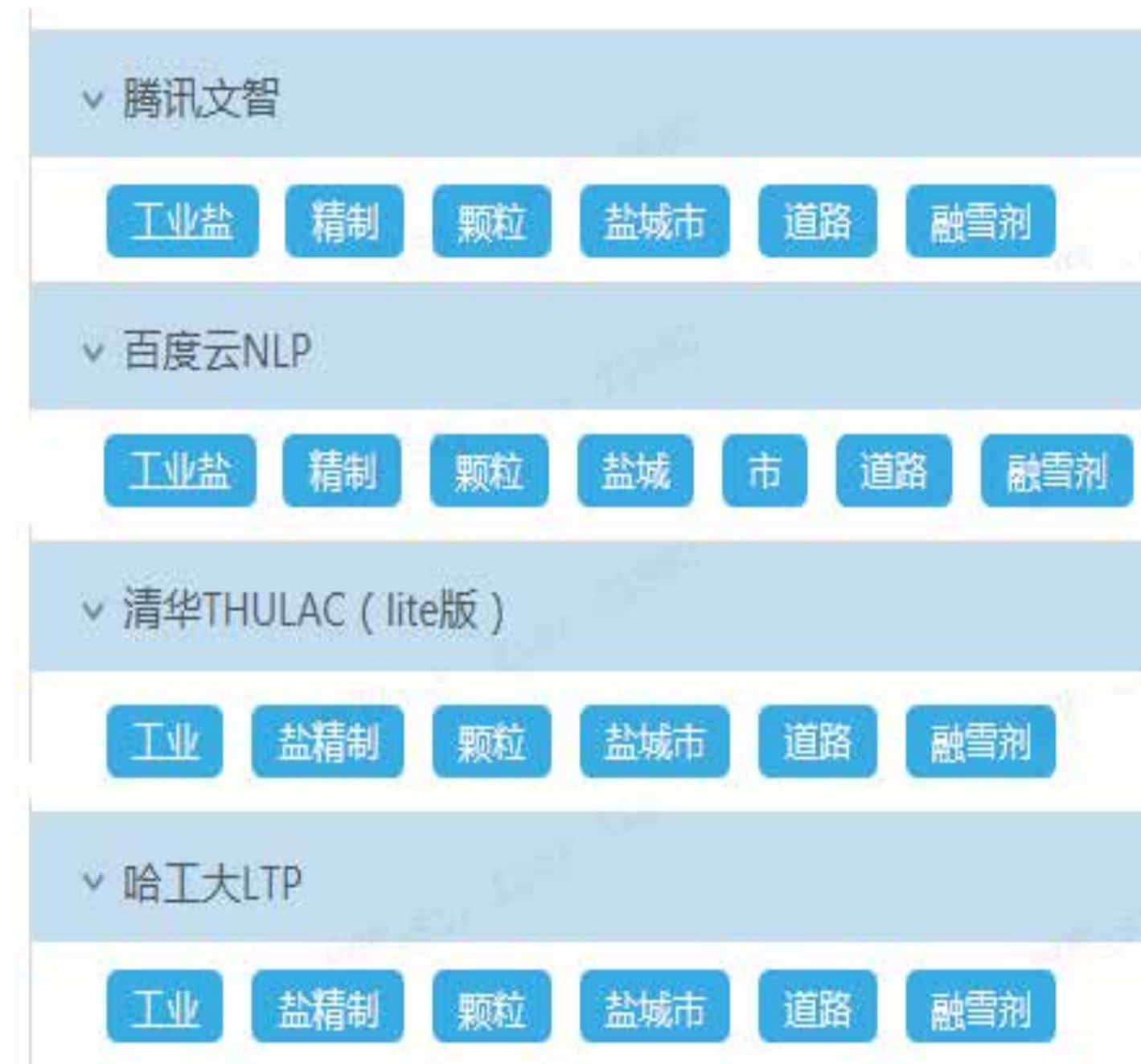
AliNLP: An NLP Platform

- Knowledge base
 - Synonyms, hypernyms, hyponyms
 - Entities, relations
- Structuralize non-structural data
 - Word segment
 - Entity extraction
 - Dependency parsing
- Sentiment analysis
- Text classification
- Question answering, chatbot
- Machine translation



AliNLP Platform

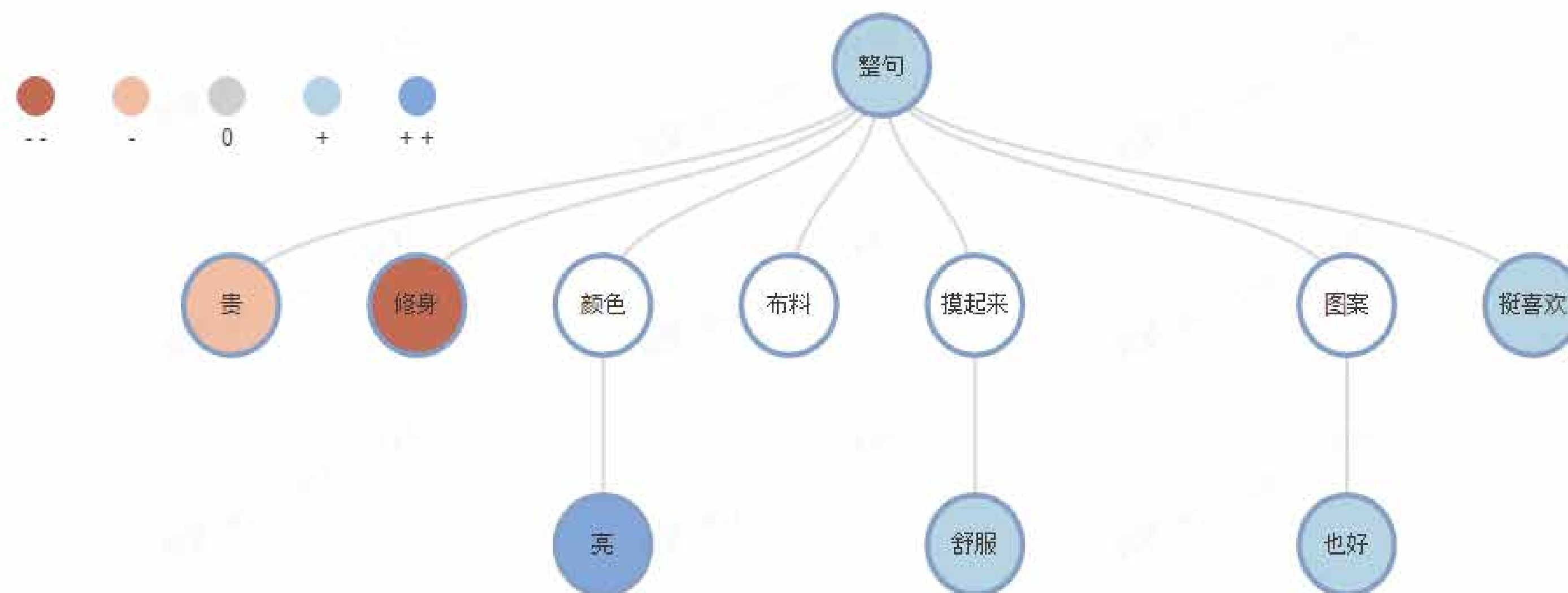
- Key component (word segmentation):
 - ✓ Our approach outperforms:
 - ✓ 1) state-of-the-art systems by 10% in E-Commerce domain;
 - ✓ 2) industrial competitors (i.e., Baidu, Tencent) by 1-2% on news data.
 - ✓ Plan: 1) Outperform the best system on academia benchmark dataset (currently 1% lower); 2) Domain adaptation model on different text genres.



AliNLP Platform

- Key component (sentiment analysis)
 - ✓ We outperform state-of-the-art systems by:
 - ✓ 5%+ in E-Commerce domain
 - ✓ 1-2% on benchmark dataset (e.g., IMDB movie reviews)
 - ✓ Plan: 1) Improve aspect-based sentiment analysis and outperform the state-of-the-art; 2) Study transfer learning to adapt to multiple domains.

虽然有点贵，不是很修身，但是颜色很亮，布料摸起来挺舒服的，图案也好看。挺喜欢的。



AliNLP Platform

- Key component (QA):
 - ✓ State-of-the-art QA relies heavily on handcrafted knowledge base. We automatically extract answers from plain text by deep learning approaches (i.e., machine reading comprehension).
 - ✓ Plan: 1) Integrate linguistic (i.e., lexical and syntactic) knowledge to deep learning models; 2) Combine NLU and user behavior models; 3) Build personalized QA

西溪园区位于文一西路969号，北靠文一西路，东临常二路，西靠高教路，南依创新路，隶属余杭区五常街道。一期项目总占地面积约16.4万平方，总建筑面积约29.3万平方，主体建筑有6幢办公楼，2幢停车楼，总工位数有14000余个，机动车位2279个，非机动车位2000多个，其中电动车充电车位820个。二期项目总占地面积约9.6万平方，总建筑面积约为15万平方，主体建筑有2幢办公楼，约5000个工位，1幢展示厅，近1号楼有中式园林的董事局办公室，将按建筑进度分期投入使用。二期预计可提供机动车位1200个。外围物业为上海明华物业管理有限公司，为园区提供安全秩序维护、设备管理维护、园区保洁等专项服务。



Q: 中式园林在哪里?
A: 近1号楼

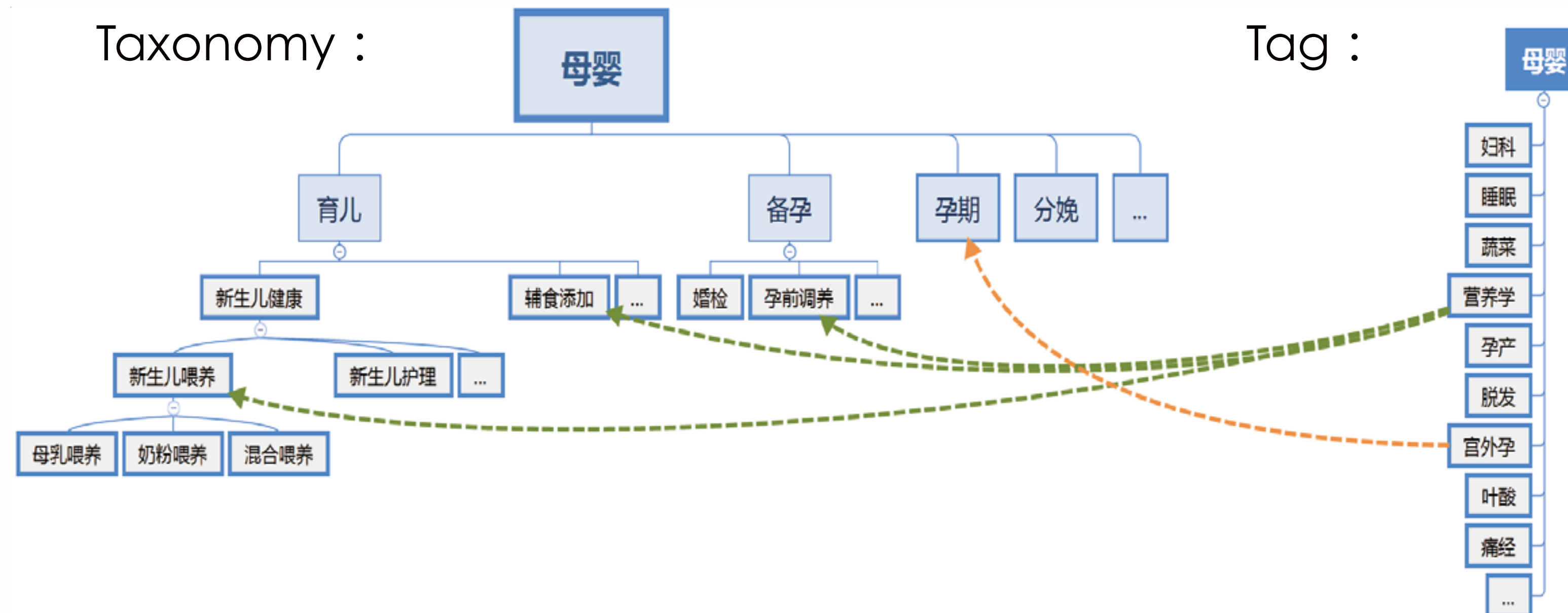
Q: 西溪园区地址?
A: 文一西路969号

Q: 西溪园区二期有多少机动车位
A: 1200

Q: 西溪园区的物业公司叫什么名字?
A: 上海明华物业管理有限公司

AliNLP Platform

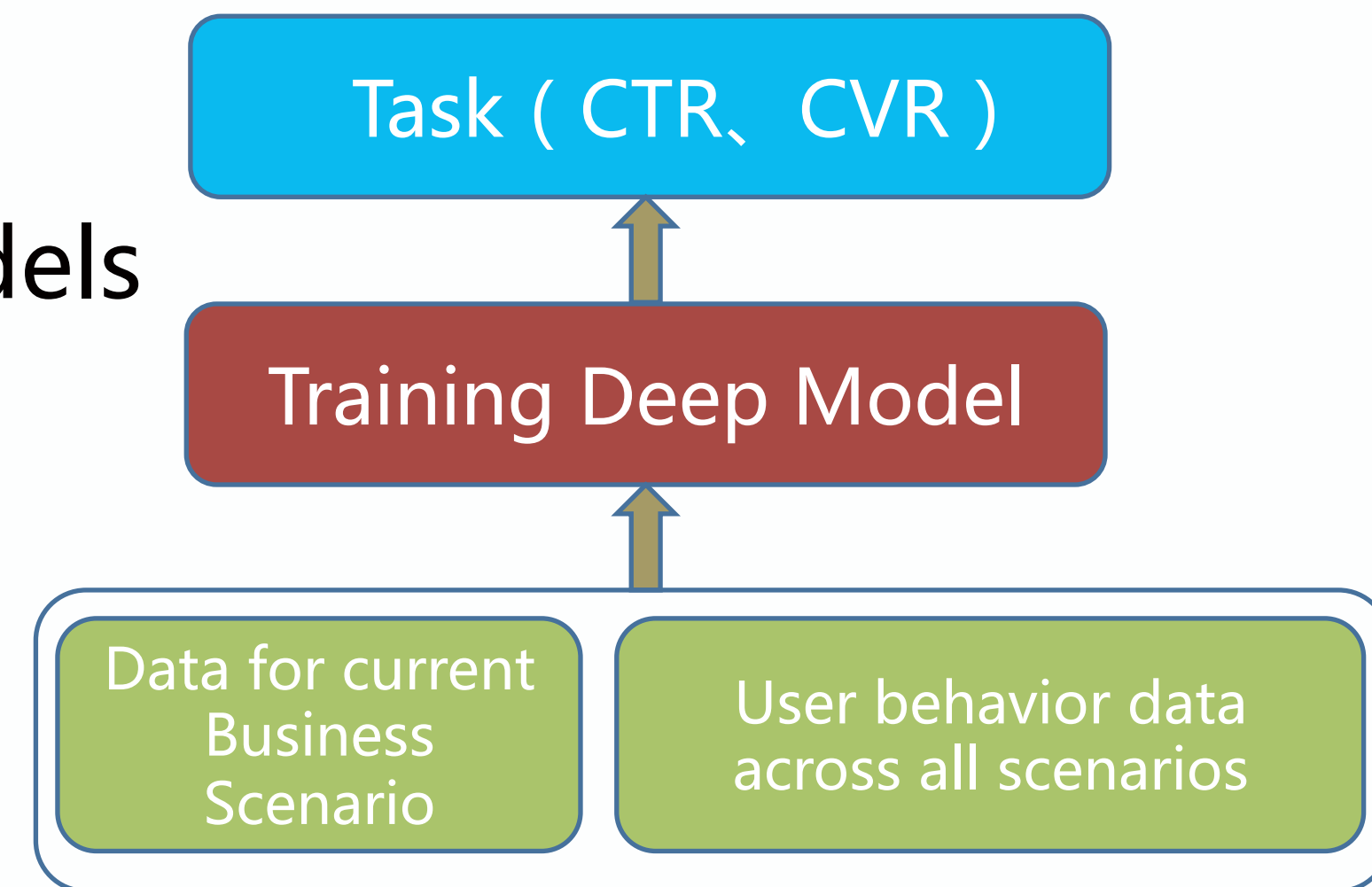
- Key component (structured content):
 - ✓ Crawler + scoring + human check, build five-level taxonomy. Use active learning to iteratively label content by tag, and associate tag to taxonomy
 - ✓ Plan: 1) Build fine-grained relations between content, tag, and taxonomy, and associate them to users; 2) Rely on unified vector representation of user and content to build user-tag Intention Graph, and detect use' s real-time intention.



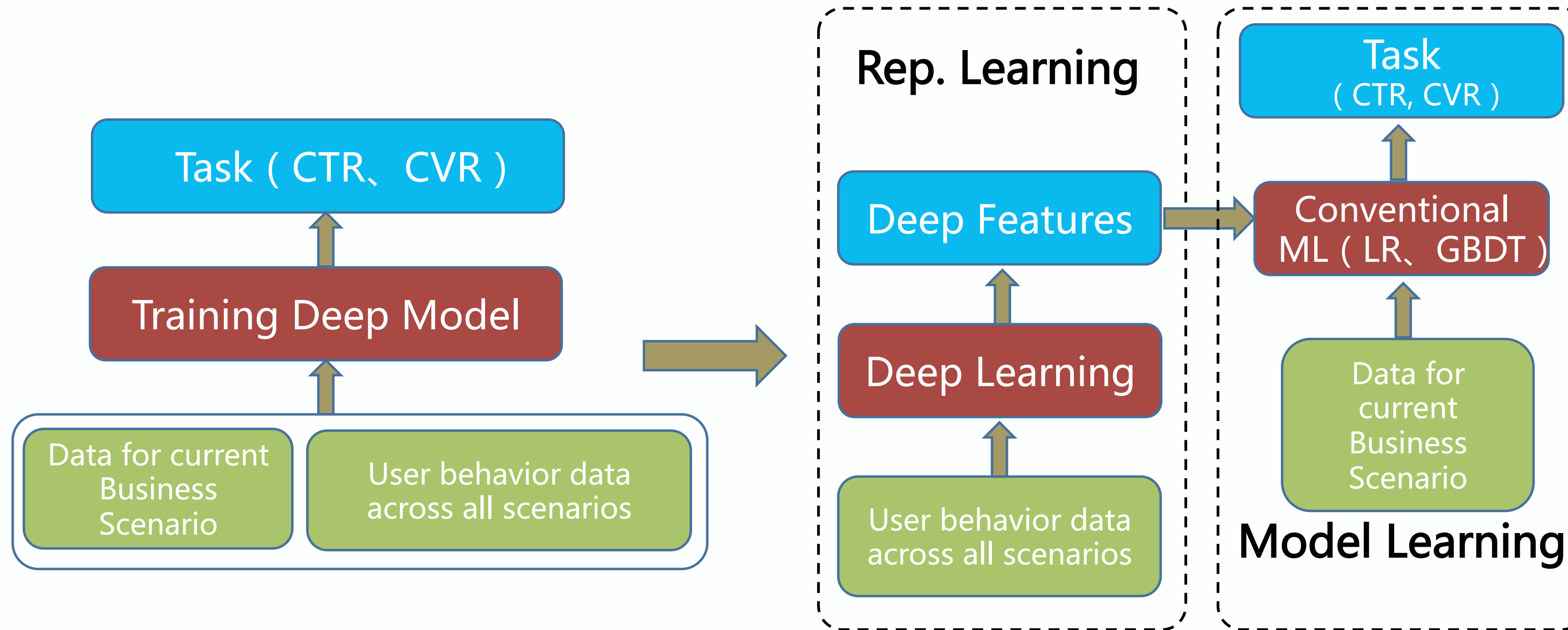
Deep Learning

Deep Representation Learning

- Need to train a deep learning model for every scenario
- Two sets of data are used in training:
 - Data for current business scenario,
 - User behavior data across all scenarios
- Drawbacks
 - High computational cost
 - Difficult to build appropriate models



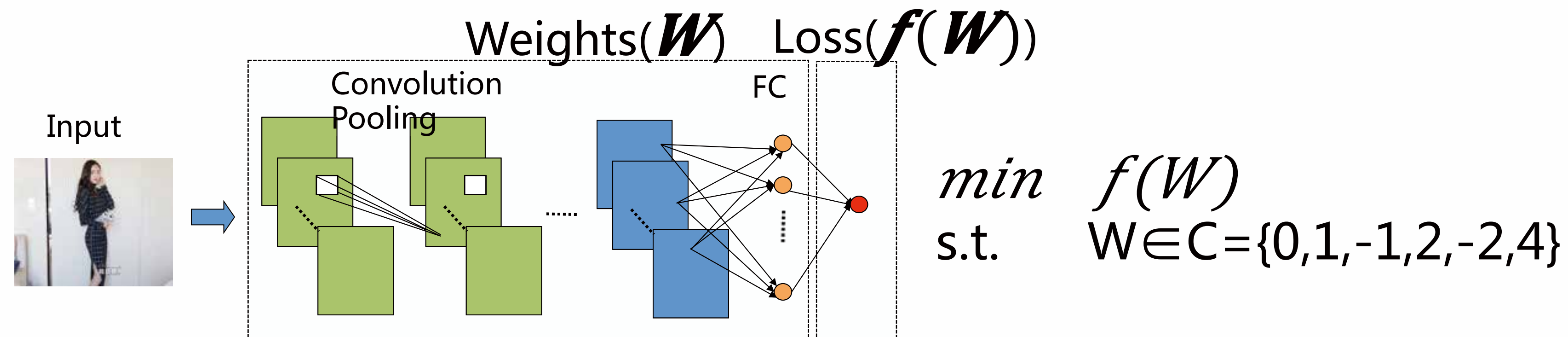
Deep Representation Learning



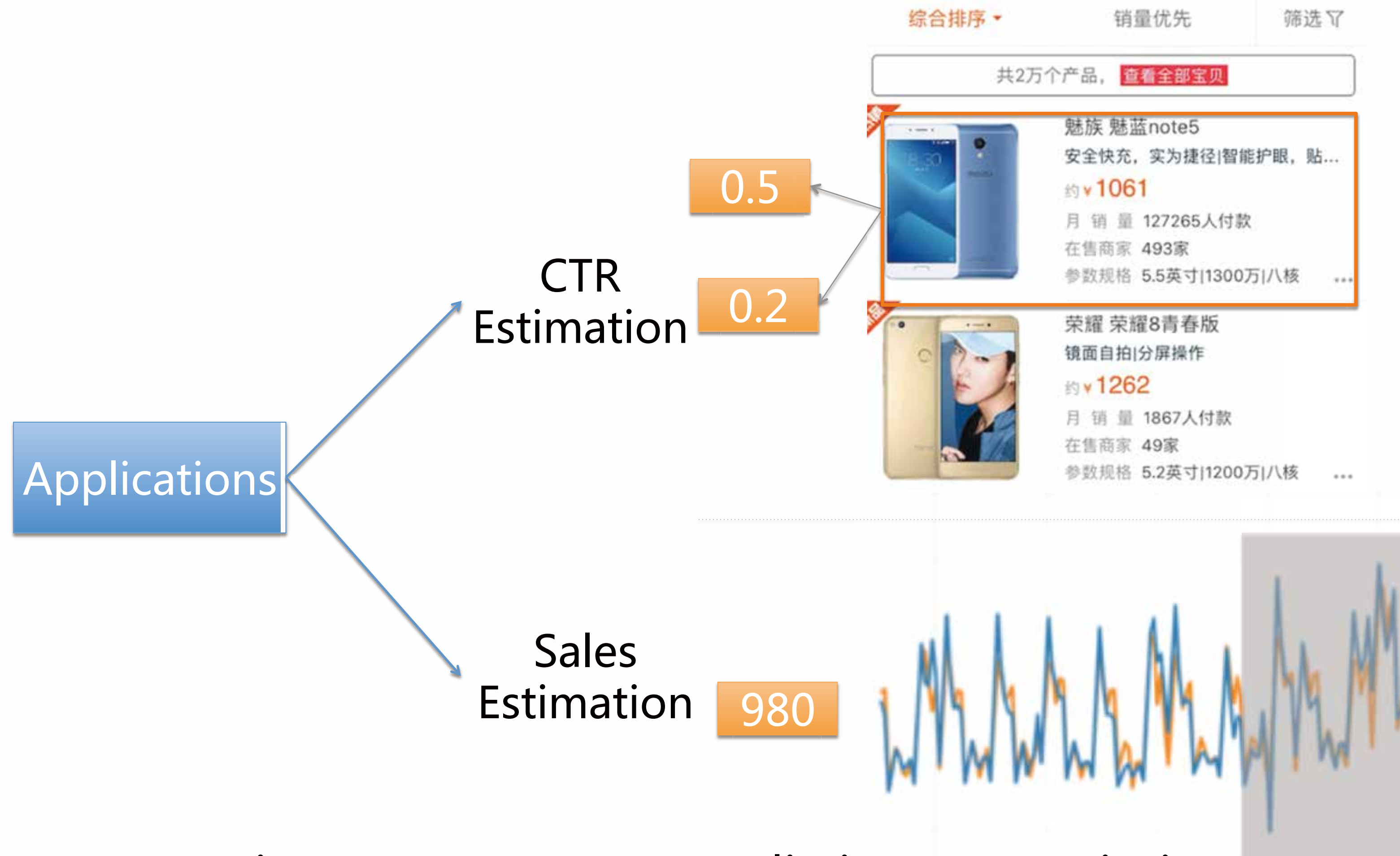
- Divide the learning task into two phases
 - Representation learning
 - Model learning
- Dramatically reduce the computational cost while maintaining similar performance

Model Compression

- Goal:
 - Reduce model size, improve computational efficiency
 - Reduce the requirement of hardware, particularly important for edge-computing
- Results
 - Reduce model size to 10% without affecting prediction performance
 - Improve prediction efficiency by 25 times on ARM/NEON



Forecasting



- Bayesian NN to capture prediction uncertainties
- Successfully applied to the single' s day (11/11) 2016

Optimization & Decision making

Optimization is Everywhere

Infrastructure



Network/Recourse Optimization



Computing Platform



Resource scheduling

 **Flink**

Machine Learning



Optimization is foundation of ML

Business



阿里数娱

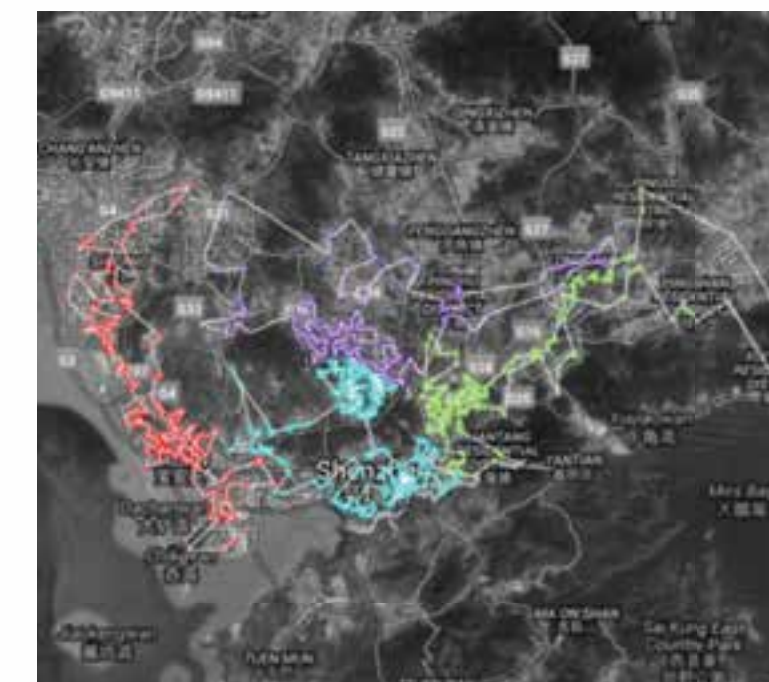
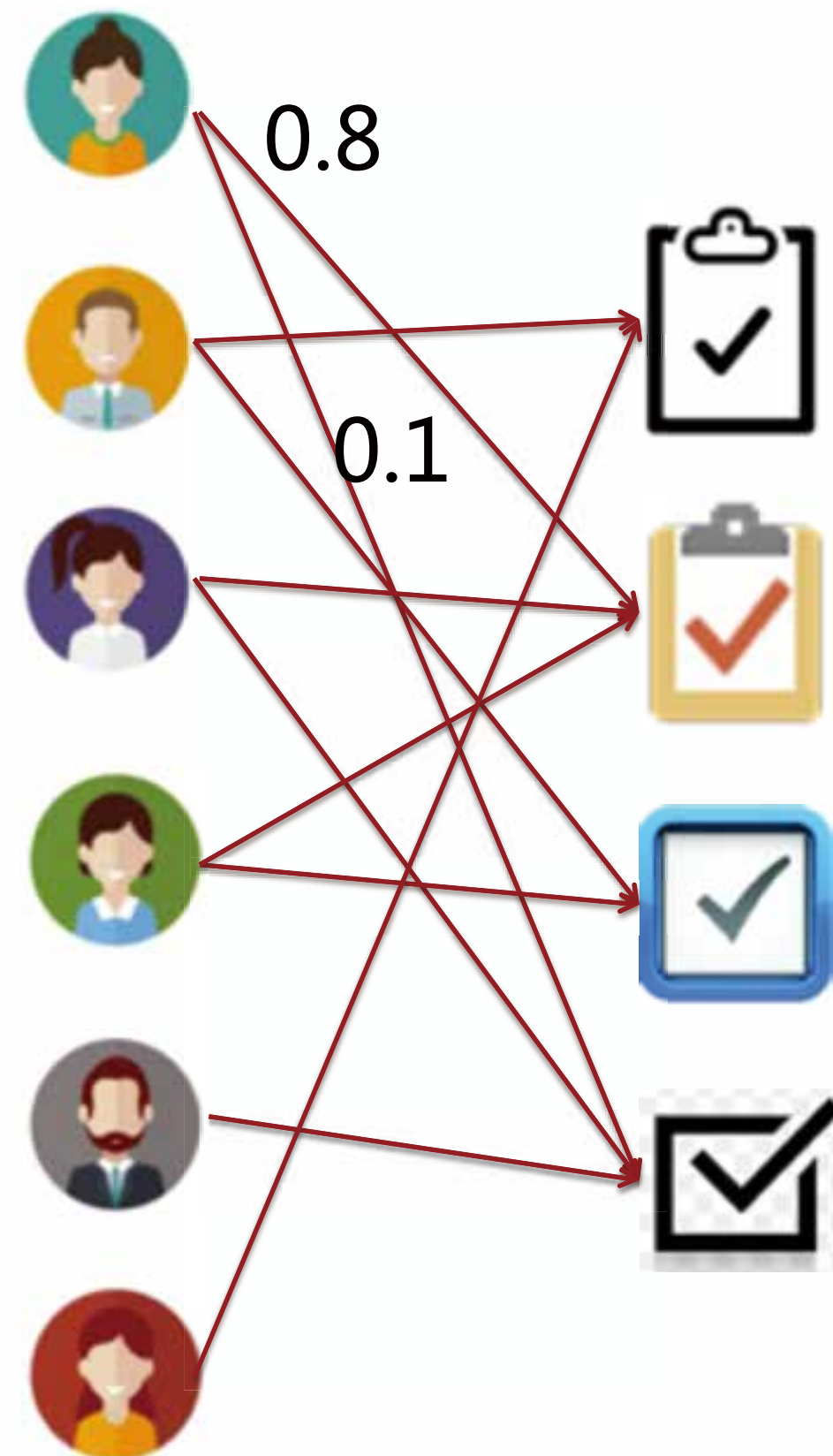
E-commerce Shipping Industry Smart City Multimedia Analysis

Optimization Engine

Assignment Prob.

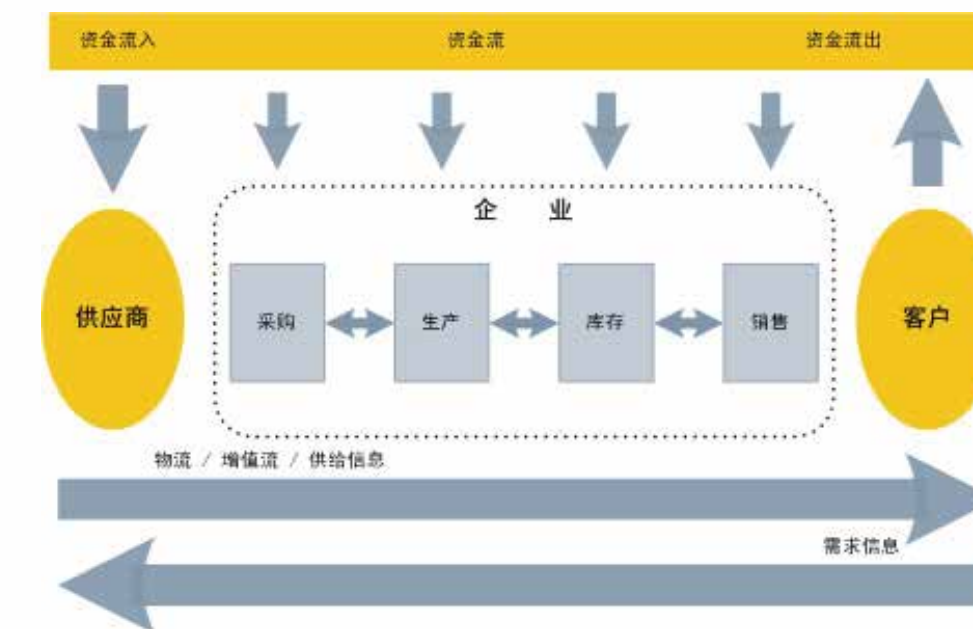
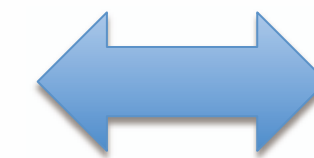
Combinatorial Opt.

Route Optimization



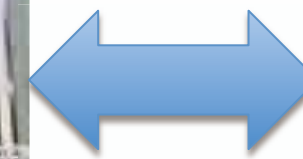
Predictive Inventory Control

- Forecasting
 - click-and-mortar, driven by AI, improved efficiency
- Decision making
 - Precision marketing
 - Automatic replenishment
 - Dynamic pricing, promotion pricing



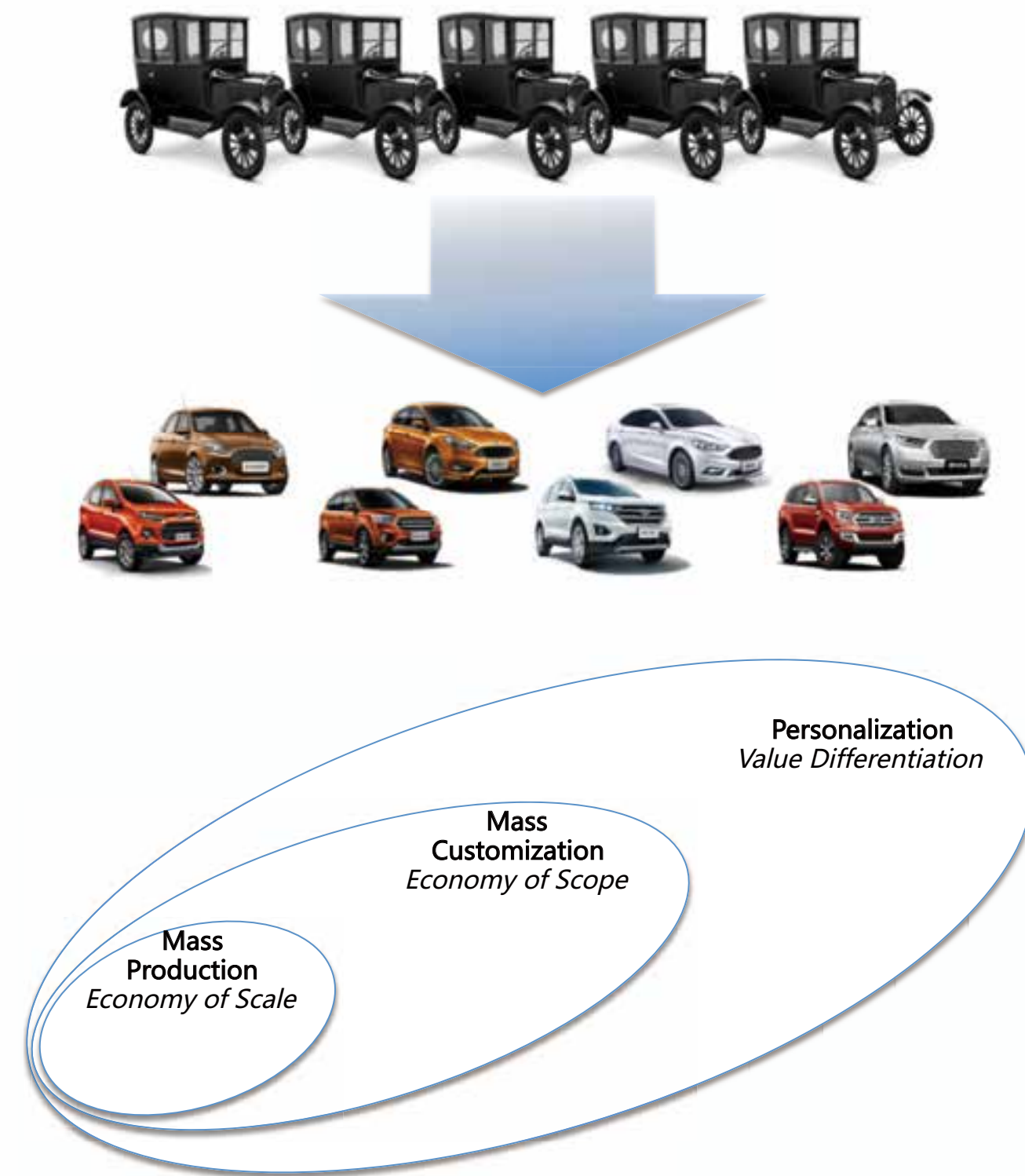
Delivery Scheduling

- Forecasting
 - Order forecasting
- Decision making
 - Delivery workforce scheduling
 - Order merging
 - Assignment
 - Routing scheduling

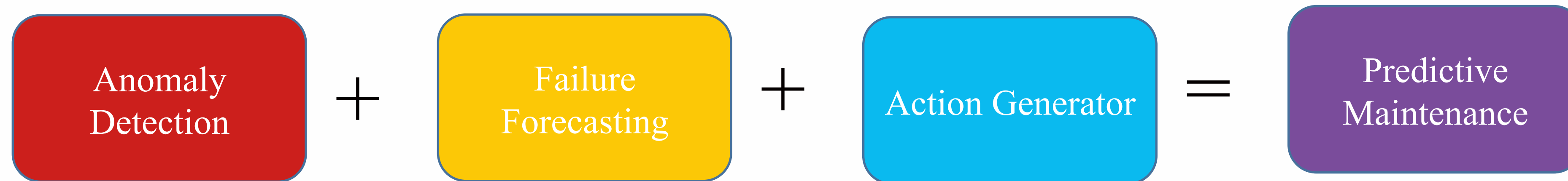


Manufacturing Scheduling

- New manufacturing
 - Shifting from mass production towards personalization (high-variety low-volume, HVLV)
- Challenges
 - Smaller granularity , shorter response time , increasing complexities and scales
- Scheduling in manufacturing
 - bridges gaps between the high-level planning and physical production
 - **Decisions** : to assign tasks onto proper resources at proper time
 - **Objectives** : to achieve production goals, such as throughput, lead-time, service level, etc.
 - **Constraints** : cost policies, due dates, human resource availabilities
- Scheduling optimization: a key to the performance improvement in HVLV manufacturing

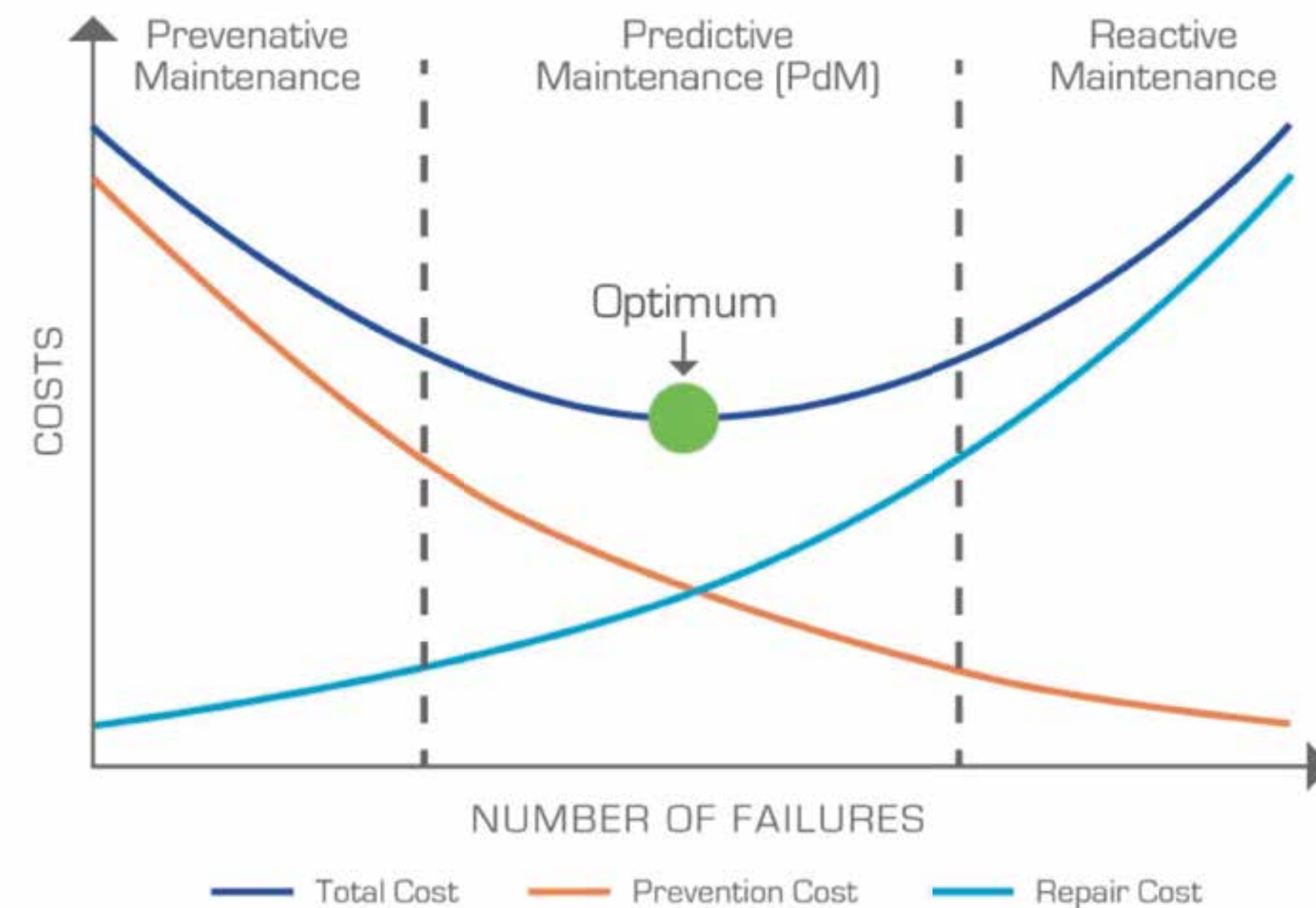


Predictive Maintenance



Predictive maintenance is the intelligent health monitoring of equipment to avoid future equipment failures

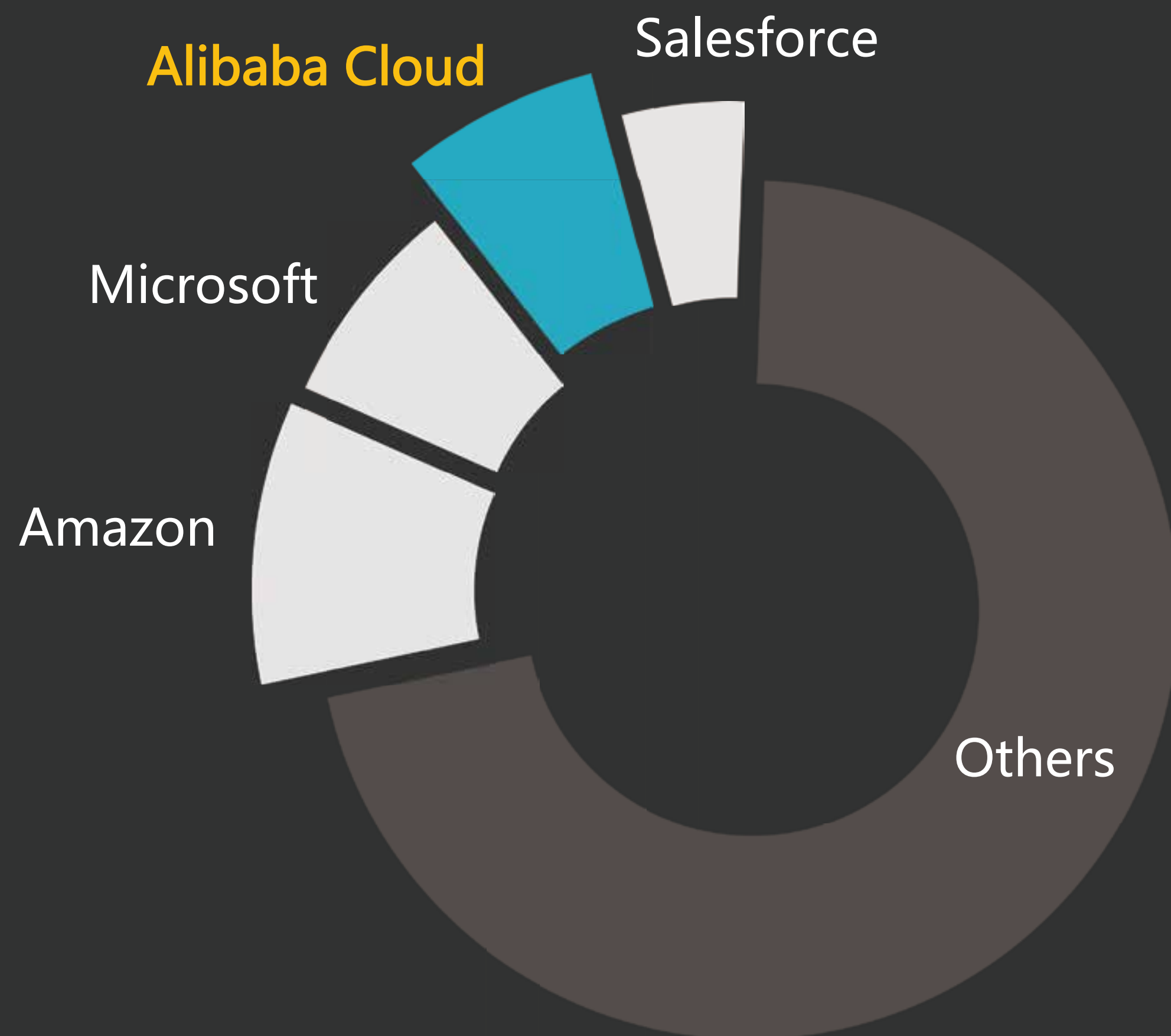
- Customized maintenance
- Reduced maintenance cost
- Higher system reliability



阿里云：为了无法计算的价值

阿里云资深专家 蔡华

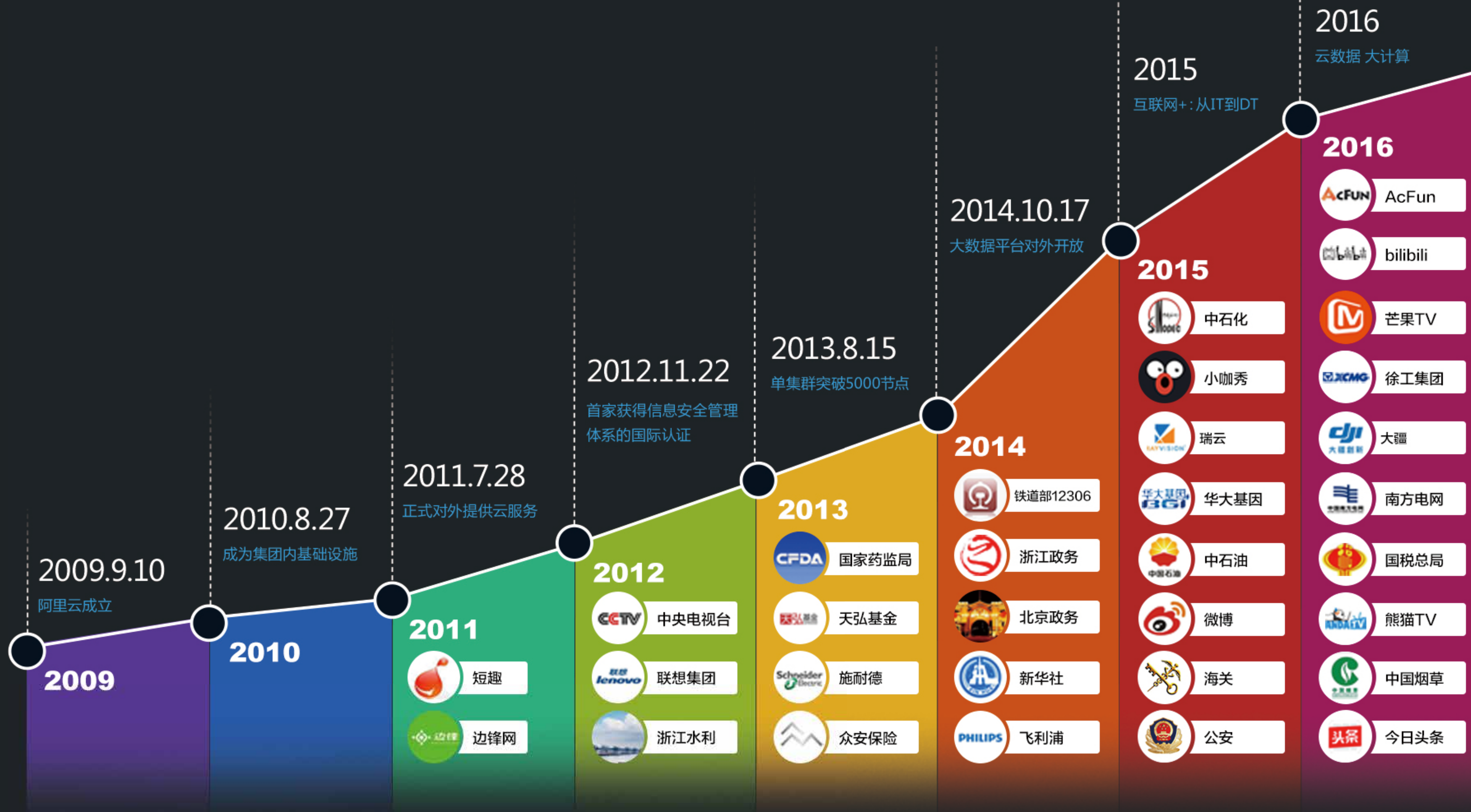
云计算全球3A格局



IDC报告显示，2015年亚太公共云市场的厂商排名阿里云位列第三。在中国本土市场获得巨大成功的阿里云在2015年跻身第三。

Source: IDC

阿里云发展历程



云计算——无处不在的算法与机器智能



Part 1

基础设施

阿里云遍布全球的基础设施能力网



无处不在的算法与机器智能



以云存储为例



- 动态规划算法用于自适应优化数据分布
- 动态子树划分算法用于管理全局名字空间
- Erasure Coding 算法用于降低数据的冗余度
- 高维聚类算法用于找出应用的负载特征
- 决策树和回归树算法用于定位问题根源
- 时间序列分析算法用于发现指标序列的异常
- 多维分析算法用于定位多因素中造成故障的原因
- 频繁项挖掘算法用于找出日志中的故障模式
- 梯度提升决策树算法、迁移学习算法用于 HDD 坏盘预测

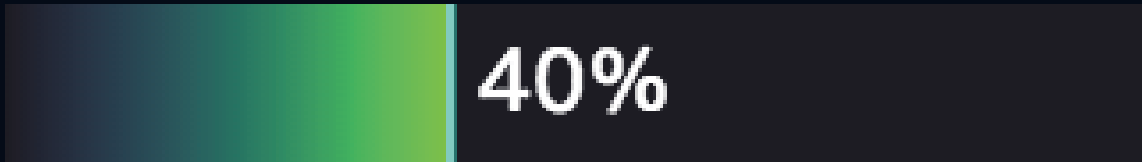
Part 2

云安全

机器智能赋能阿里云安全

超大规模的基础防御能力

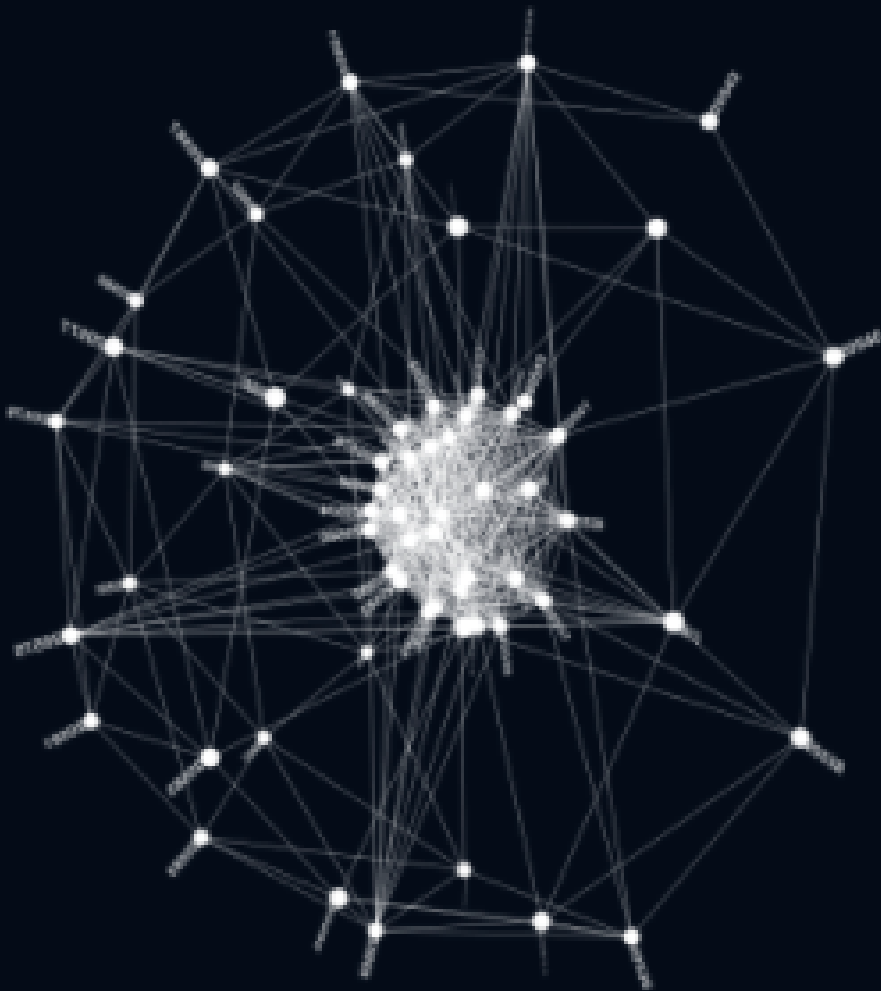
保护全国40%网站



今天抵御攻击 1,600,809,900次



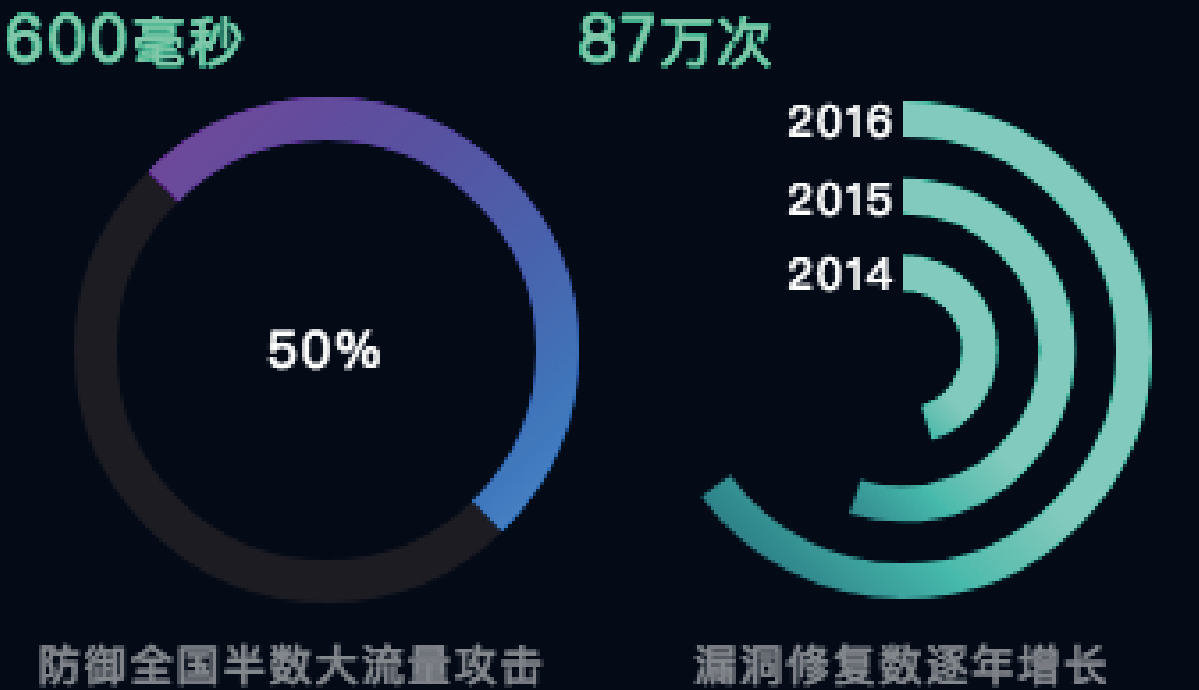
平台威胁感知神经网络



运行133个威胁检测模型
每日分析60TB数据, 严重告警1分钟响应



阻断云平台DDoS攻击 2016年帮助用户修复漏洞



国际攻击源

1 美国 United States	4463
2 巴西 Brazil	1688
3 印度 India	711
4 韩国 Korea	690
5 俄罗斯 Russia	655

全球攻击态势

攻击源位置	攻击源IP	攻击类型	防御状态
意大利	164.132.xx.xx	web攻击	拦截成功
中国	113.195.xx.xx	暴力破解	拦截成功
中国	221.229.xx.xx	端口扫描	拦截成功
保加利亚	194.141.xx.xx	暴力破解	拦截成功
中国	121.201.xx.xx	暴力破解	拦截成功
中国	210.44.xx.xx	暴力破解	拦截成功
香港	103.213.xx.xx	web攻击	拦截成功

云安全中的算法与机器智能

云平台安全

暴力破解检测分类——线上实时精准拦截，拦截量提升2倍以上，无误拦

网页风险内容检测NLP——对色情、赌博、涉政等保持95%以上的检测精确率，30%的违规做到算法自动检测并处罚

主机入侵检测——去除80%以上的误报警

云安全产品 - 云盾

机器学习Web应用防火墙——提升规则引擎30%的精确率，无误报，每笔检测在5毫秒内完成

反爬虫防刷单——抑制黄牛刷单抢座，支持亚洲航空和12306春节售票

4层7层CC攻击检测——CC攻击的唯一有效保护伞

恶意二进制文件深度学习检测

前沿研究

深度学习及其安全应用

GAN

强化学习

Part 3

ET大脑

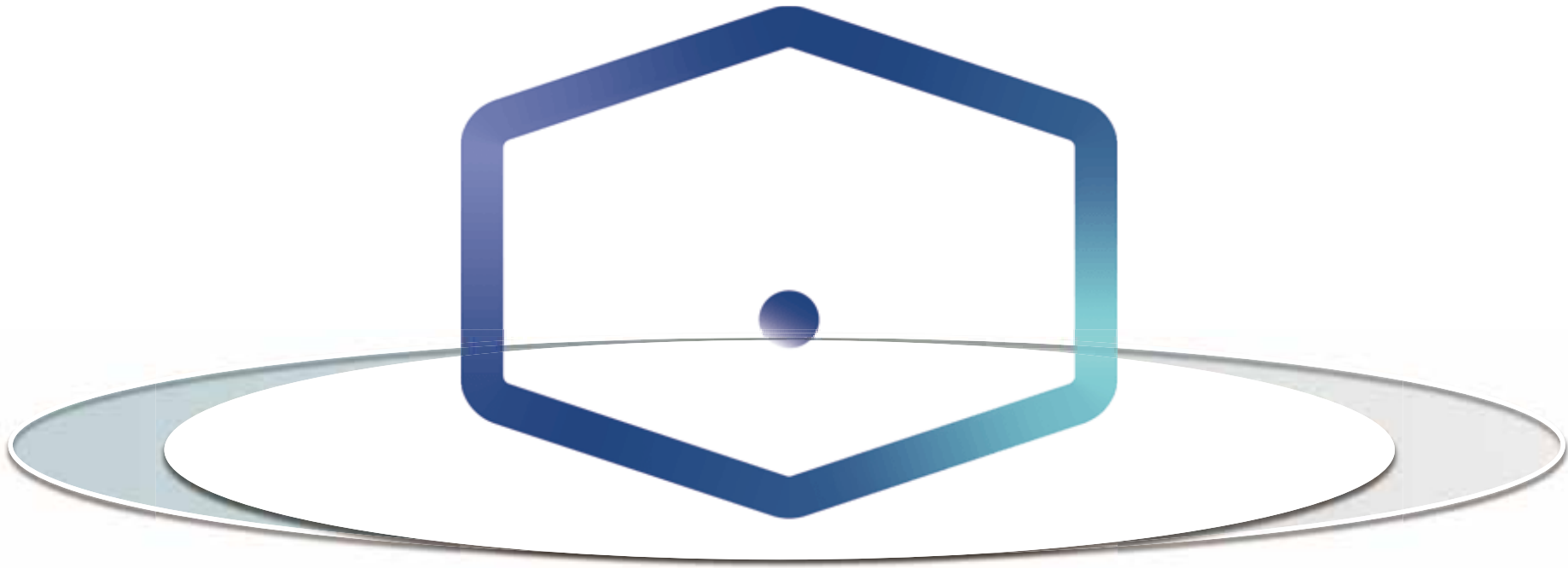
ET大脑的能力

多维感知、全局洞察、实时决策、持续进化在复杂局面下快速做出最优决定

认知
多维感知



决策
全局洞察



ET BRAIN

ET 大脑

判断
实时决策



学习
持续进化

ET大脑

杭州大脑

试点区域通行时间减少15.3%
高架路出行时间节省4.6分钟

萧山大脑

特种车辆到达时间缩短50%

苏州大脑

公交优化试点线路乘坐率提升17%

首都机场

廊桥停机位利用率提高10%，相当于每天有20000名旅客不用再乘坐摆渡车

协鑫光伏

通过对太阳能电池切片制造流程进行最优参数推荐，提升了1%的切片良品率

中策橡胶

优化密炼工艺：门尼值标准差降低14%、密炼时长减少10%，密炼温度降低6%，炼胶合格率提升5%

盾安风电

提前1~2周识别风机潜在故障，变大修为小修，大幅降低维修成本，缩短维修时间

天合光能

通过对天合光能生产关键环节的优化，帮助天合光能实现了7%的A品比例提升

城市大脑开放创新平台——国家四大人工智能平台之一
城市大脑永久入驻国家博物馆

ET大脑

数据智能应用



工业大脑



城市大脑



金融大脑



环境大脑



医疗大脑

行业最佳实践

政府

金融

视频

能源

互联网

交通

电商

医疗

工业

航空

...

计算平台



离线计算



图计算



分析型数据库



搜索引擎



流计算



EMR

数据采集和
数据同步

数据加工和分析工具



数仓开发套件



QuickBI



机器学习



数据可视化



关系网络分析



公众趋势分析



画像分析

人工智能和引擎



智能语音交互



人脸识别



推荐引擎



文字识别



图像与视频分析



营销引擎



规则引擎



调度引擎

Part 4

视频云

用技术构建不一样的视频云



视频云算法

客户端

低功耗视频编解码	可编排视频流水线	视频滤镜、特效	物体跟踪
高清视频采集/播放	统一多媒体框架	网络自适应推流	AR特效
GPU/CPU混合视频处理	网络自适应播放	非线性视频编辑	VR视频
短延时互动直播	组件化多媒体引擎	人脸识别	客户端AI处理

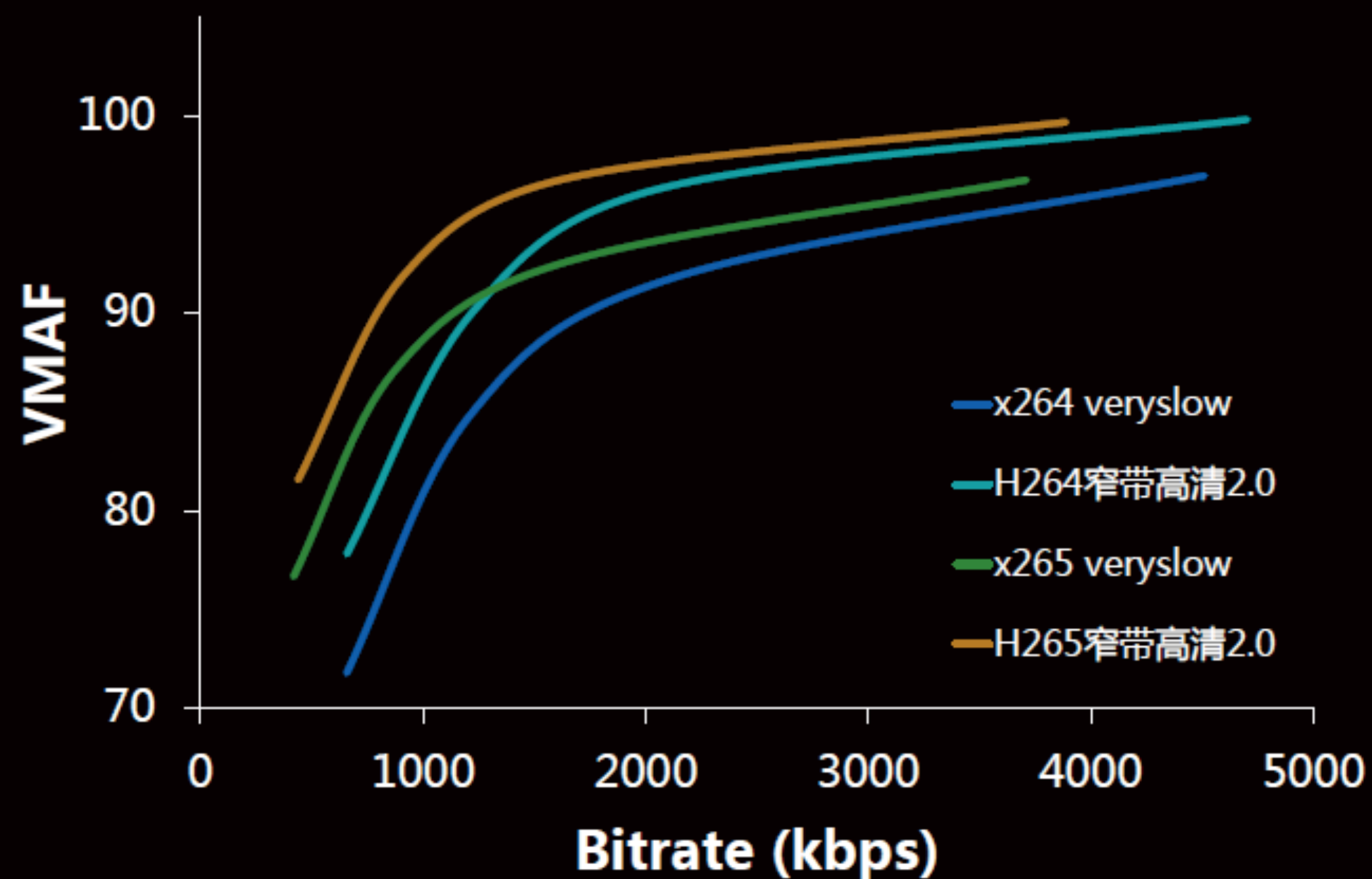
服务端

海量弹性视频处理	画质增强、重生	低延时互动视频	视频加密及DRM
实时视频质量监控	超高清视频处理	HEVC编解码	全球视频传输
分布式、实时转码	窄带高清转码	视频资源防盗	智能视频内容理解
多模态音视频分析	视频指纹	智能视频审核	视频智能编辑

窄带高清2.0

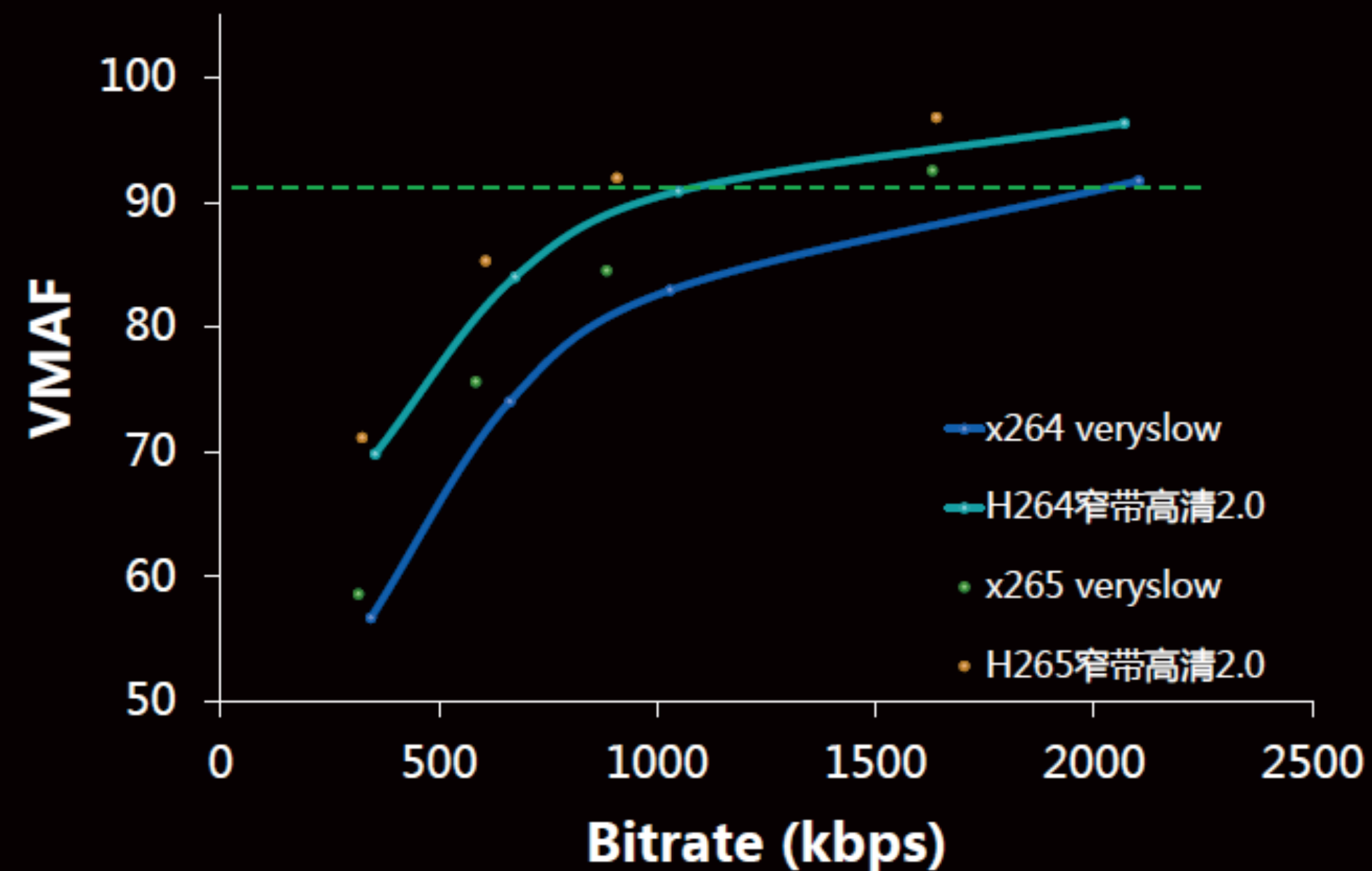
让H.264达到H.265的效果，让720P达到1080P的效果

1080p编码的RD曲线 (VMAF)



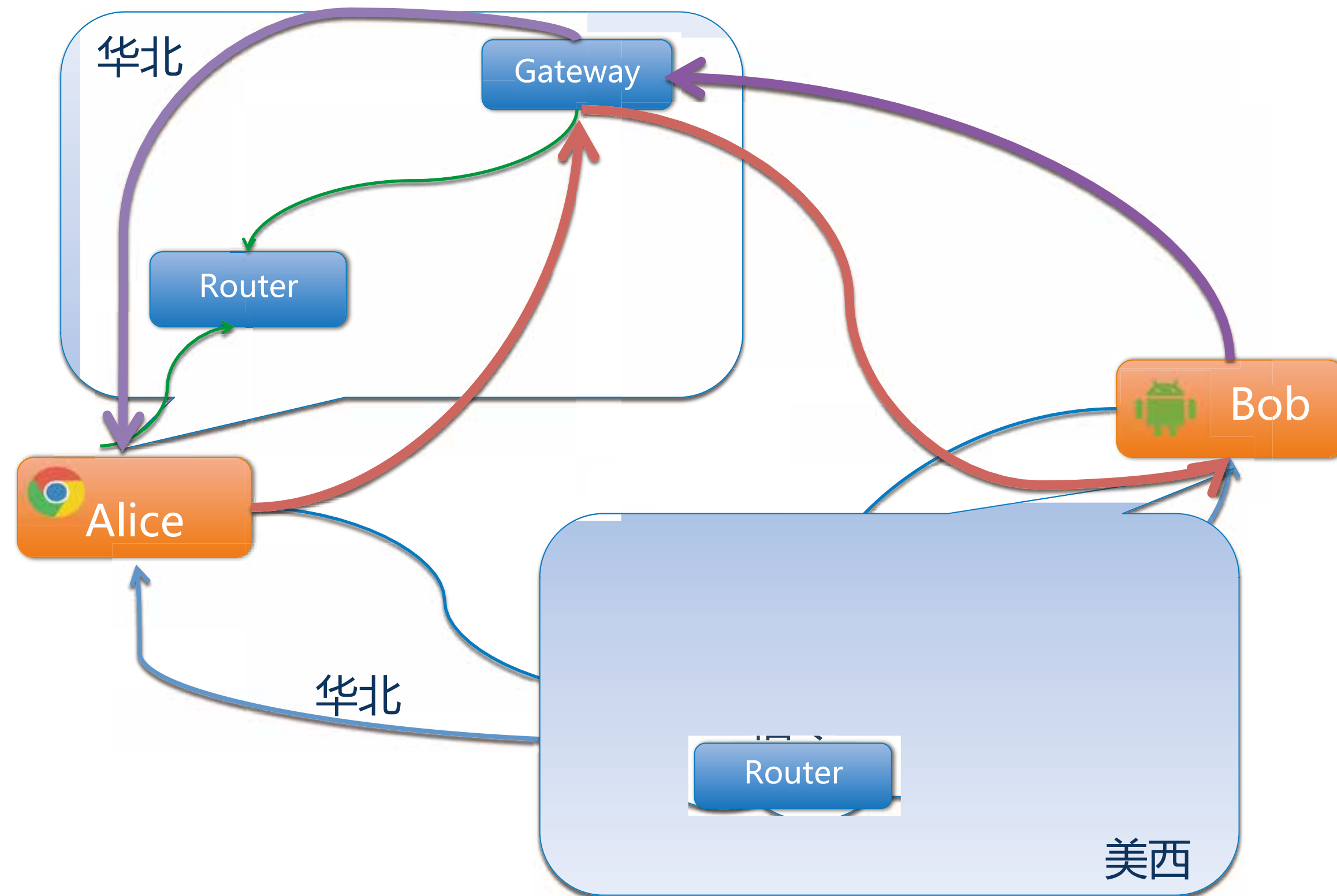
✓ 超越编码器能力上限，提前进入下一代水准

跨清晰度的RD曲线 (VMAF)



✓ 超越分辨率观感上限，低分辨率也能体验高清

实时音视频通信



全球超低延时组网

社交，会议，教育，娱乐，医疗和其他
数以千计的网络节点
高清音视频，安全可靠

实时视频编解码与传输

信源信道联合优化的弱网对抗
基于主观视觉的质量最优
自适应质量和性能调整

网络后台

全网质量监控和智能调度
基于大数据和机器学习的优化
下一代传输协议

视频编解码算法



VP8



VP9

H.266



Alliance for
Open Media



webp

HEIF

丰富的应用场景

直播，点播，短视频

音视频通信（SVC）

静态图片压缩

视觉质量优化

客观质量评估

基于主观质量的编码器优化

基于深度学习的前后处理

速度与功耗优化

快速算法：FME，FMD.....

异构计算：CPU，GPU，FPGA

汇编指令集：Neon，SIMD

下一代标准研究



H.266



Web  RTC



视频编解码标准

H.266 , AV1 , AVS

提出多种应用场景的痛点和需求
技术创新，推动下一代标准

音视频通信标准

WebRTC 2.0的需求和提案

其他相关标准

Video Quality Expert Group
IETF Internet Video Codec
行业联盟



为了无法计算的价值

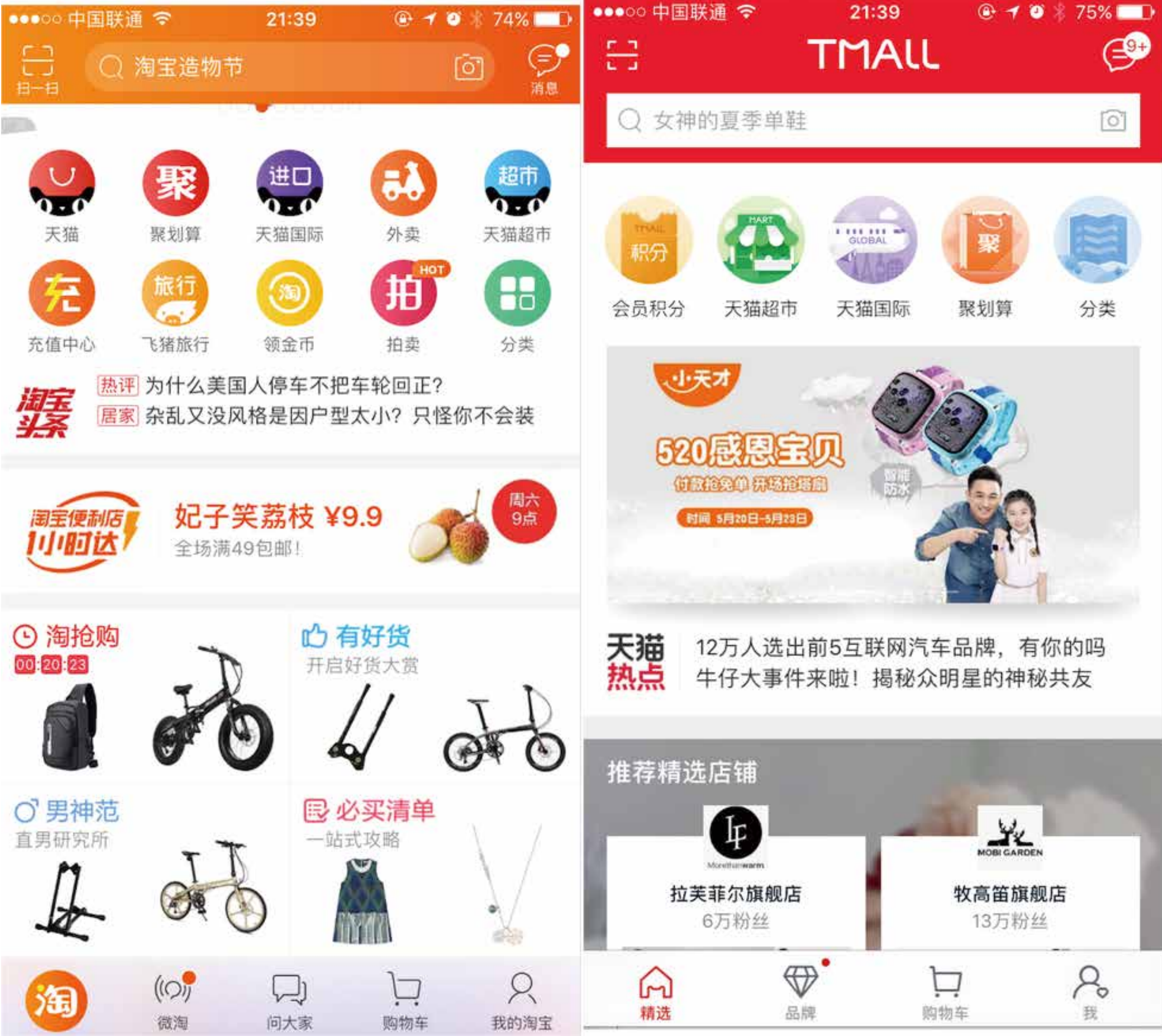
走进淘宝技术部算法世界

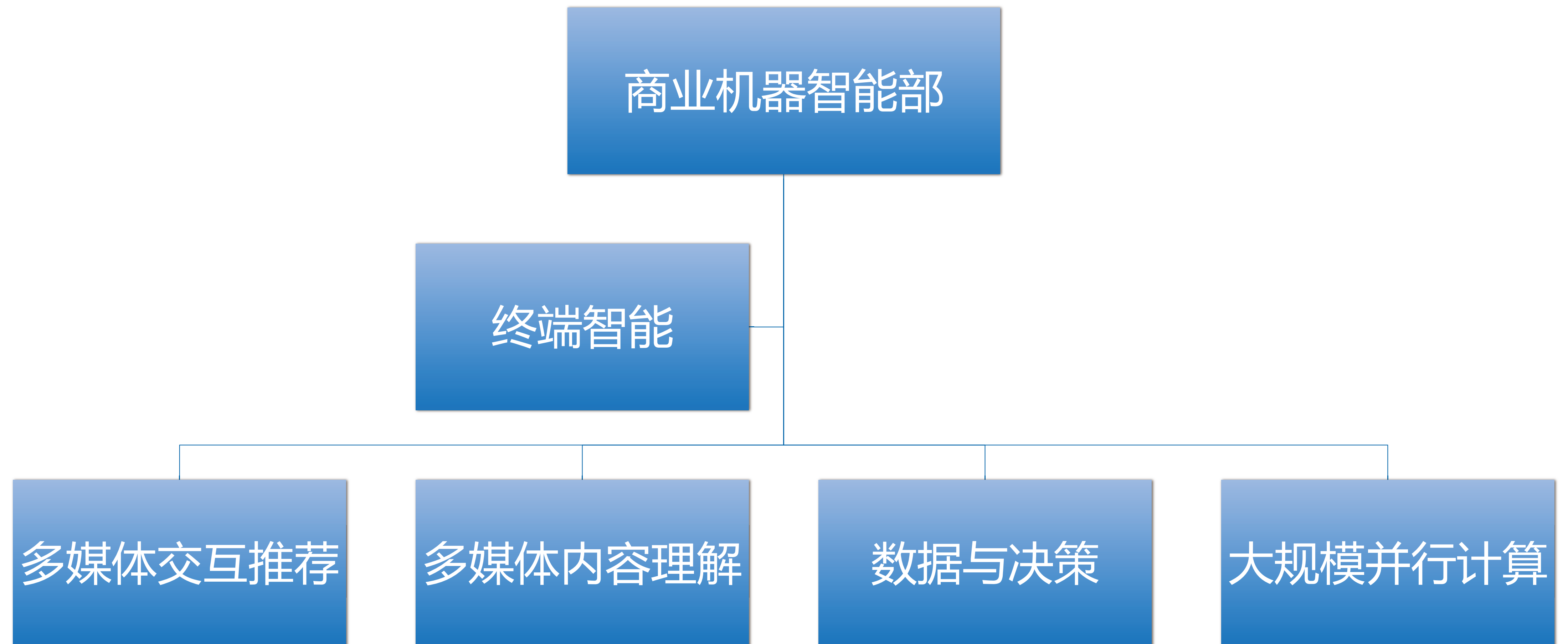
商业机器智能部资深算法 永叔

部门介绍

淘宝技术部是一支具有光荣历史和深度互联网技术影响力、具备全球领先技术实力的综合性团队，涉及分布式计算、音视频通信、数字媒体、机器学习、视觉算法、IoT。我们的愿景是致力于成为全球最懂用户和商业的技术创新团队，目标是服务10亿消费者，以科技打造一个更加简单友好的商业世界。

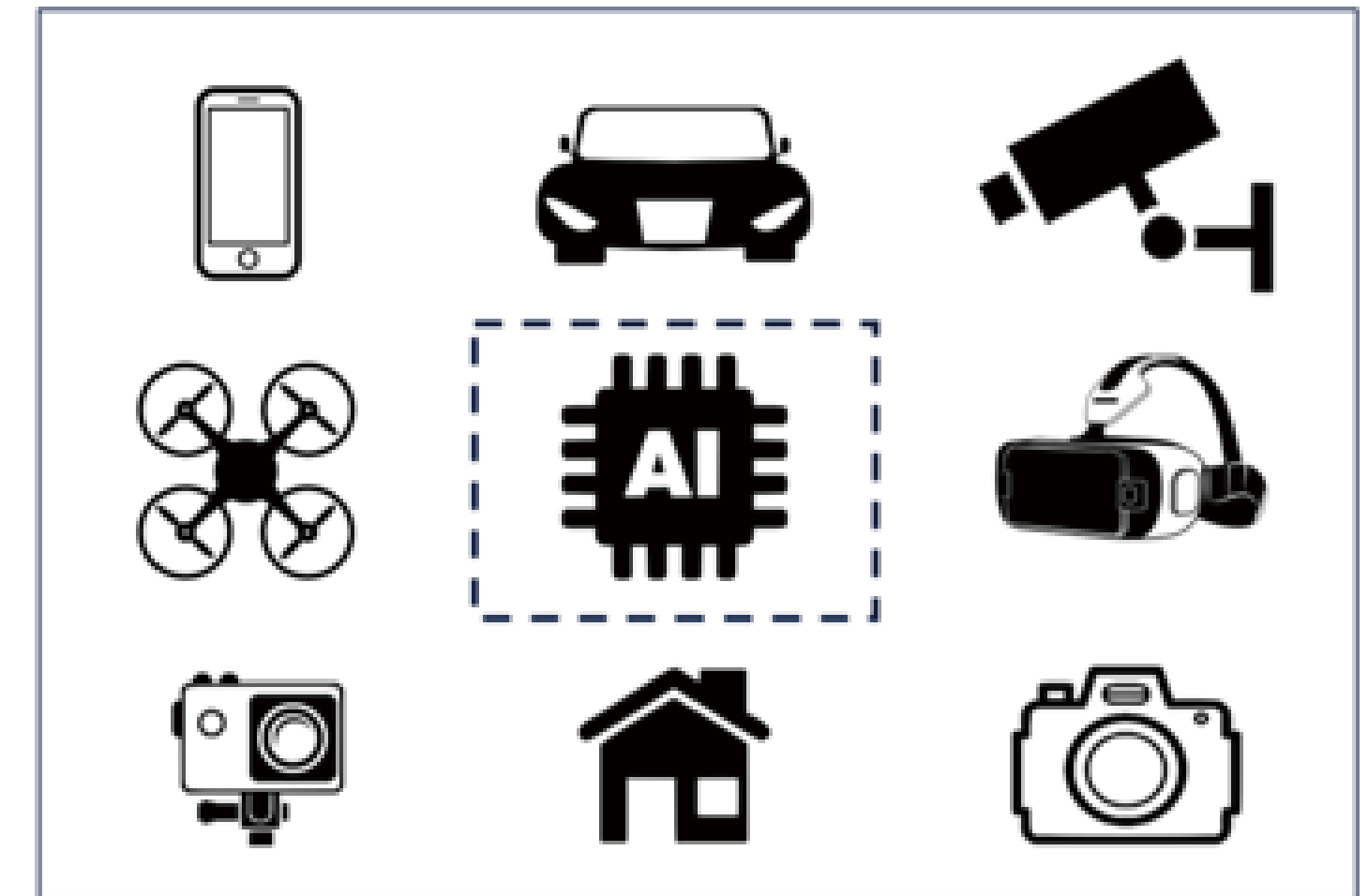
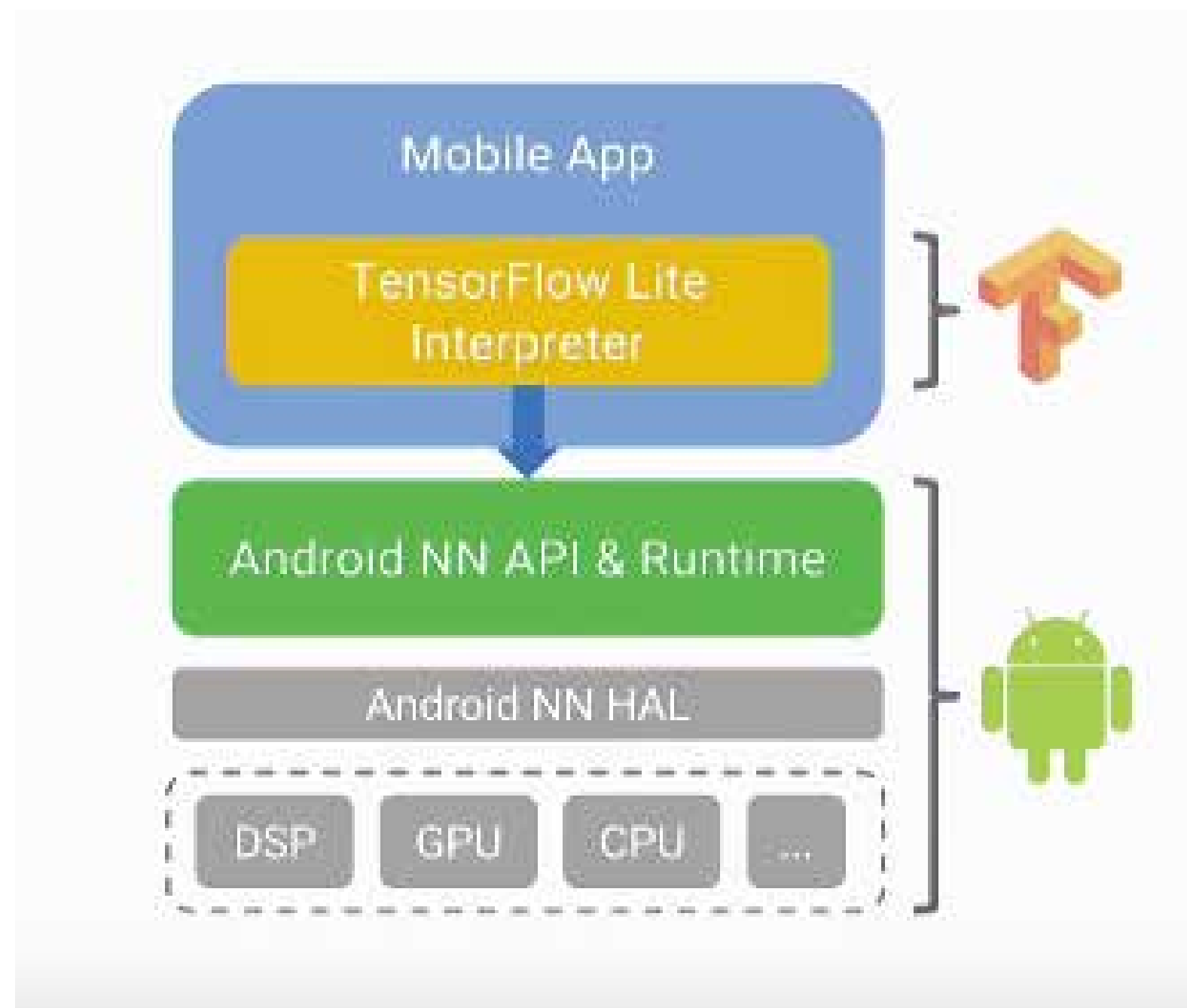
业务版图





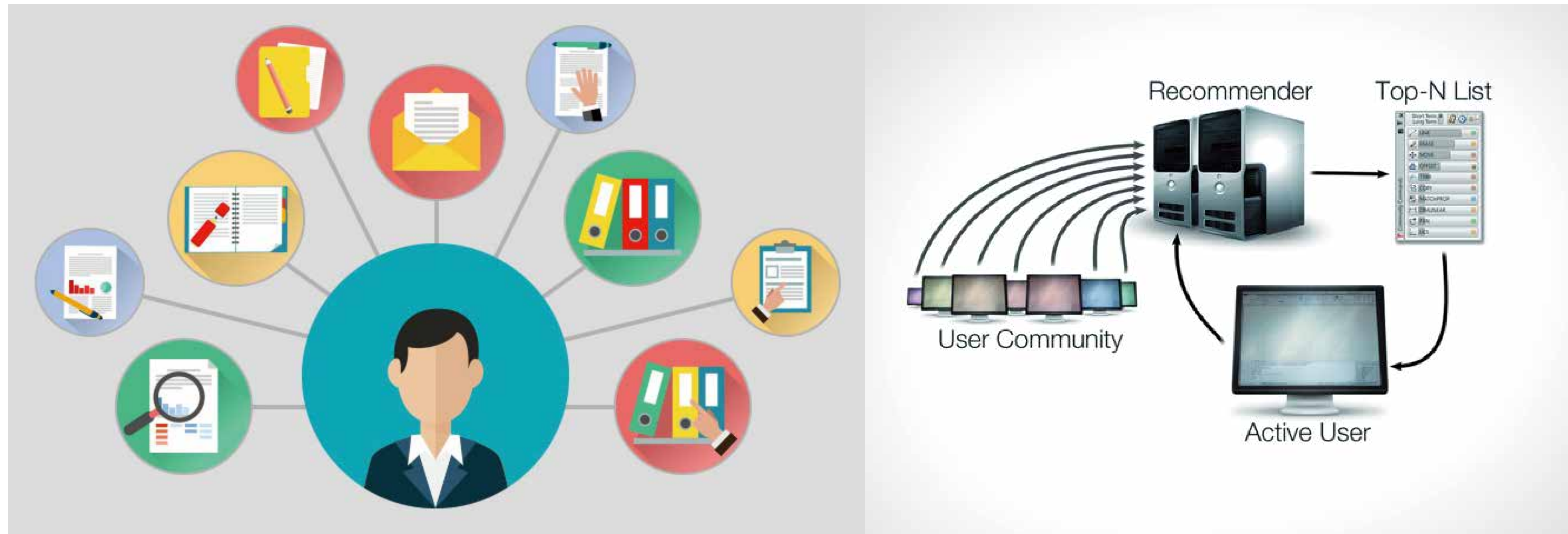
移动AI&终端智能

云+端为用户带去更为实时和智能化的新体验



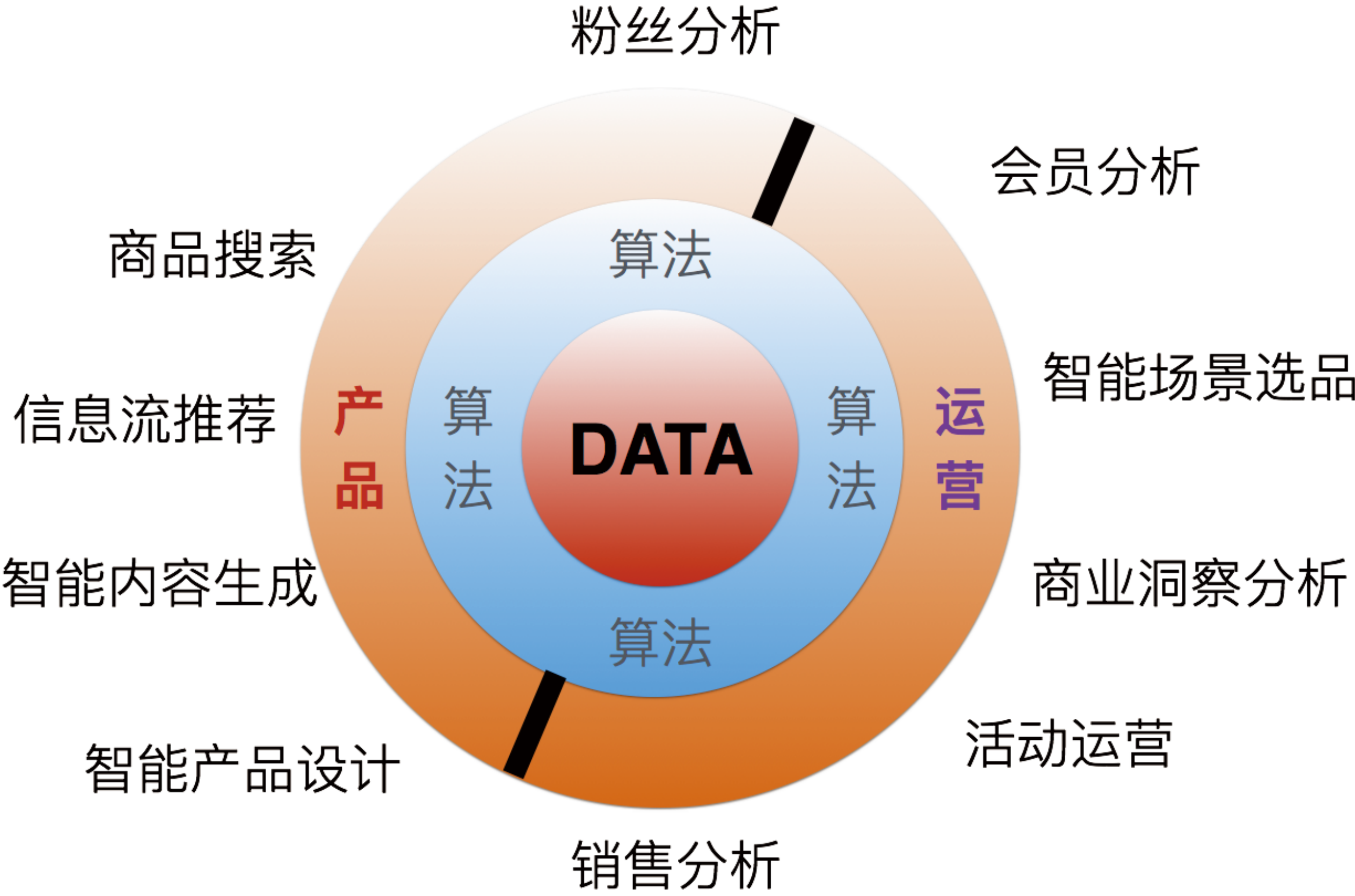
多媒体交互式推荐

综合利用语音、视频、智能对话机器人等技术打造下一代融入式交互推荐系统



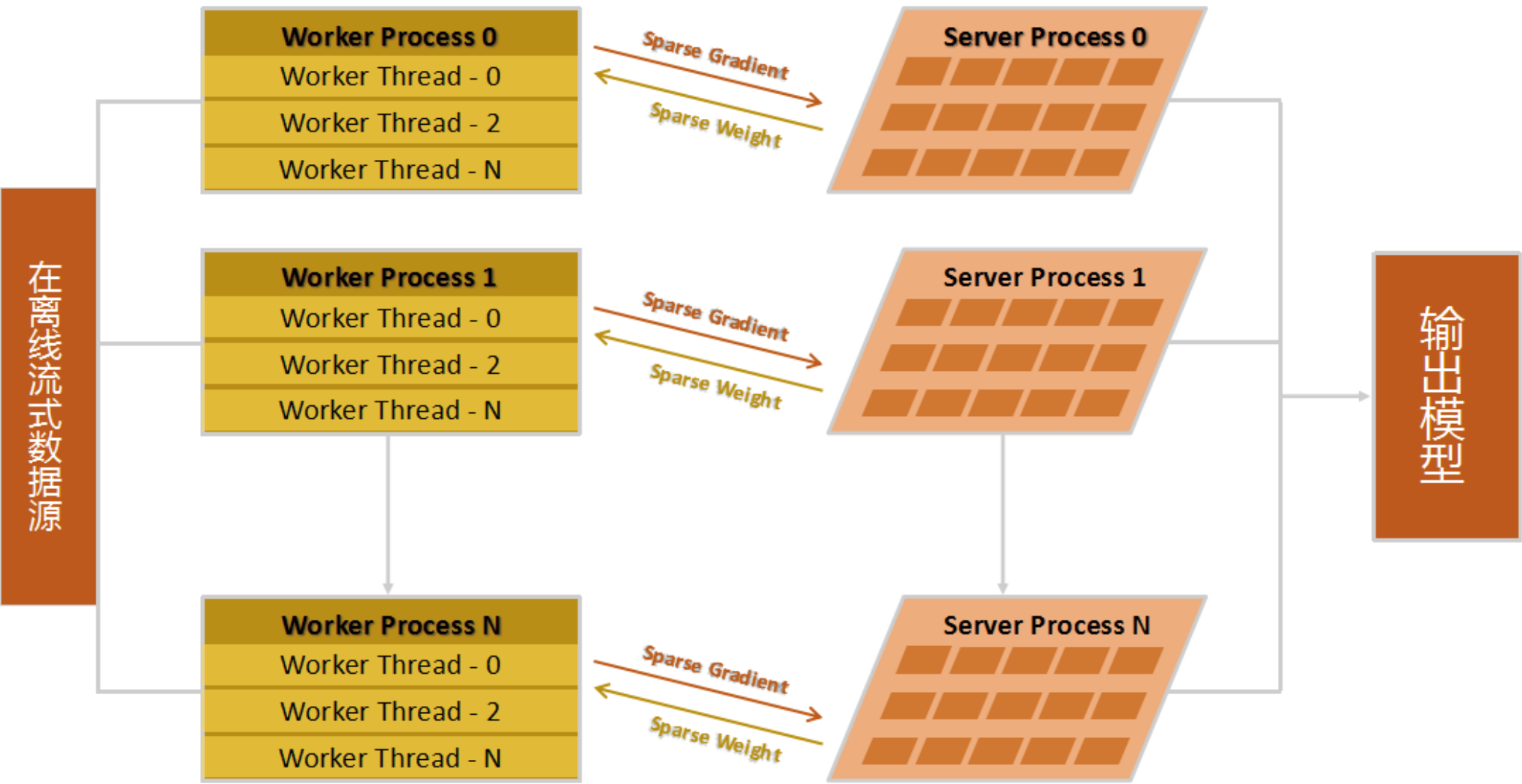
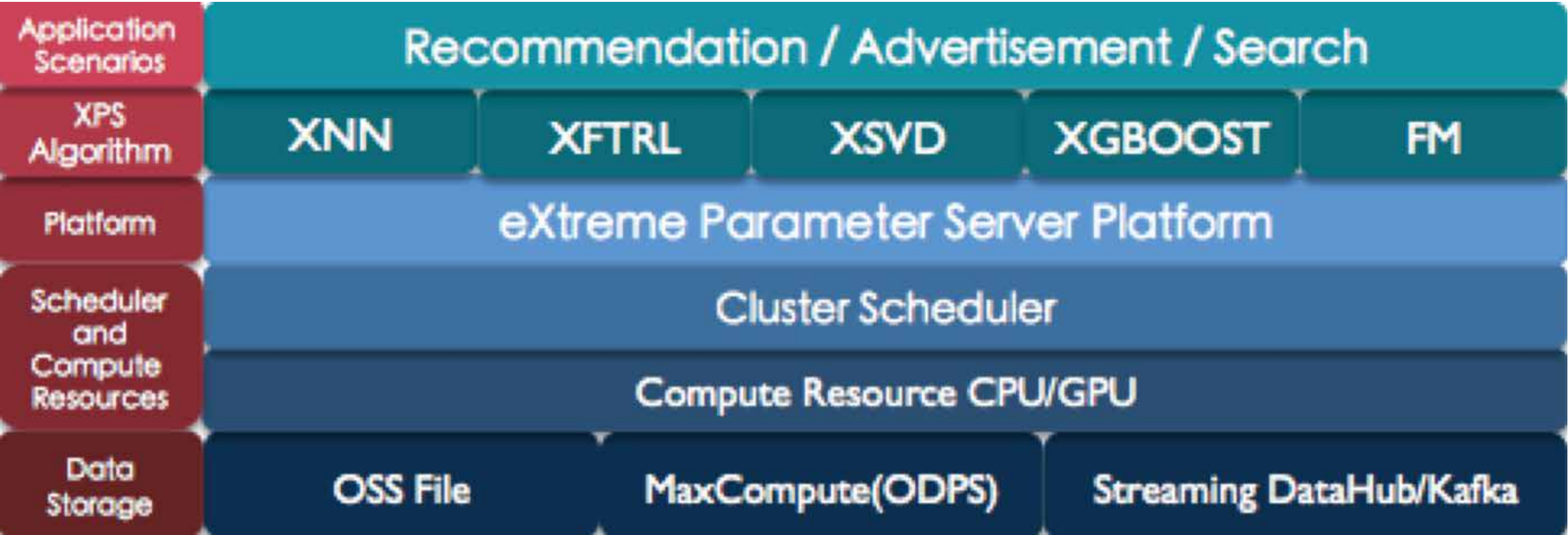
数据与决策

为新零售商家及行业运营的每一次营销提供智能决策支持



大规模并行计算

面对每天超过百亿的训练数据，我们持续在研发更快、效果更好的分布式机器学习算法平台及系统



商业机器学习部需要的人才



机器学习算法工程师

图像/视觉算法工程师

自然语言处理算法工程师

数据挖掘算法工程师

图像和美

- 2009年成立，阿里巴巴时间最久的图像组
 - 专注于商品图片的算法研究和应用
- 构建了集团内应用最广泛的图像基础设施
 - ✓ 千亿级图像搜索引擎
 - ✓ 文字识别
 - ✓ 服饰智能搭配

图像和美团队介绍

Vision & Beauty Team

时尚之心 since 2011

- 前沿AI技术结合时尚专家知识，独有的AI技术/产品体系
- 服务用户日常穿衣场景，并提供专属搭配建议
- 落地线下智能体验店，引领服饰新零售潮流
- FashionAI国际挑战赛，发布业界首个同时满足服饰专业性和机器学习要求的AI挑战平台



http://fashionai.alibaba.com/?spm=a2c22.188588.992154.20.4e1845d09YiSWy&lang=zh_CN#home

图像和美团队介绍

Vision & Beauty Team

图像搜索与识别 since 2009

- 千亿级实时图像搜索引擎
- 为阿里巴巴构造了非标准类目的“同款”概念以及数据
- 为50万卖家提供了图像知识版权保护系统



我们需要

- 对技术创新充满激情的你，投入到以下工作中：
 - 时尚相关商品图像的细粒度属性识别以及关键部位定位；
 - 自动化的商品图像组织与呈现方式研究；
 - 基于拍照图片的商品图像分析；
 - 基于拍照图片的用户脸型身形分析；
 - 基于NLP的穿搭理由生成；

图像和美团队介绍

Vision & Beauty Team

文字识别 since 2010

- 负责阿里巴巴集团的绝大部分OCR需求，日处理量6亿+；
- 商品图详情图识别准确率接近人眼（97.6%），耗时实时（200ms）；
- 覆盖了几百种表单等数据类型；
- 云产品“读光”将对外发发布，结合行业知识，形成行业信息化解决方案。
- 联合华南理工大学共同举办ICPR MTWI 2018 挑战赛；



<https://baijiahao.baidu.com/s?id=1581058399005148577&wfr=spider&for=pc>

我们需要:

人工智能领域的一流人才投入到以下工作中：

1. OCR识别算法研发；
2. 表格，文本理解，知识抽取等自然语言处理工作；
3. 跟踪业界进展，技术交流以及专利申请，论文发表；
4. 深度学习理论研究以及优化



天猫技术：电商“最强大脑”

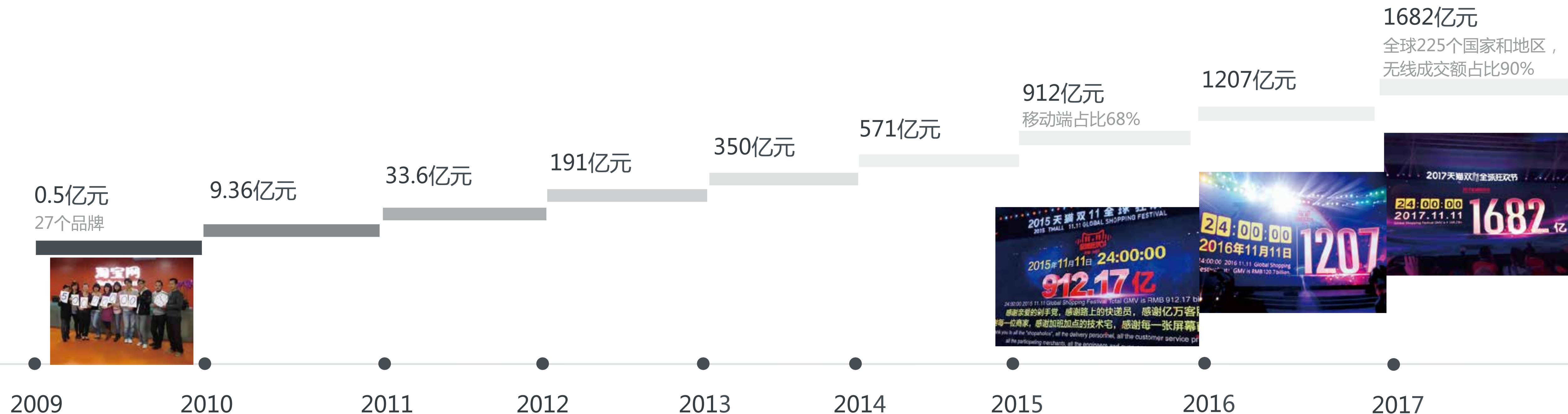
营销平台资深算法专家 志昭

目录



认识天猫，从双11开始

天猫作为全球最大的B2C购物平台，每年双11都在不断刷新全球单日最大购物成交记录



全球消费者挚爱的品质购物之城

文化：客户第一，突破创新

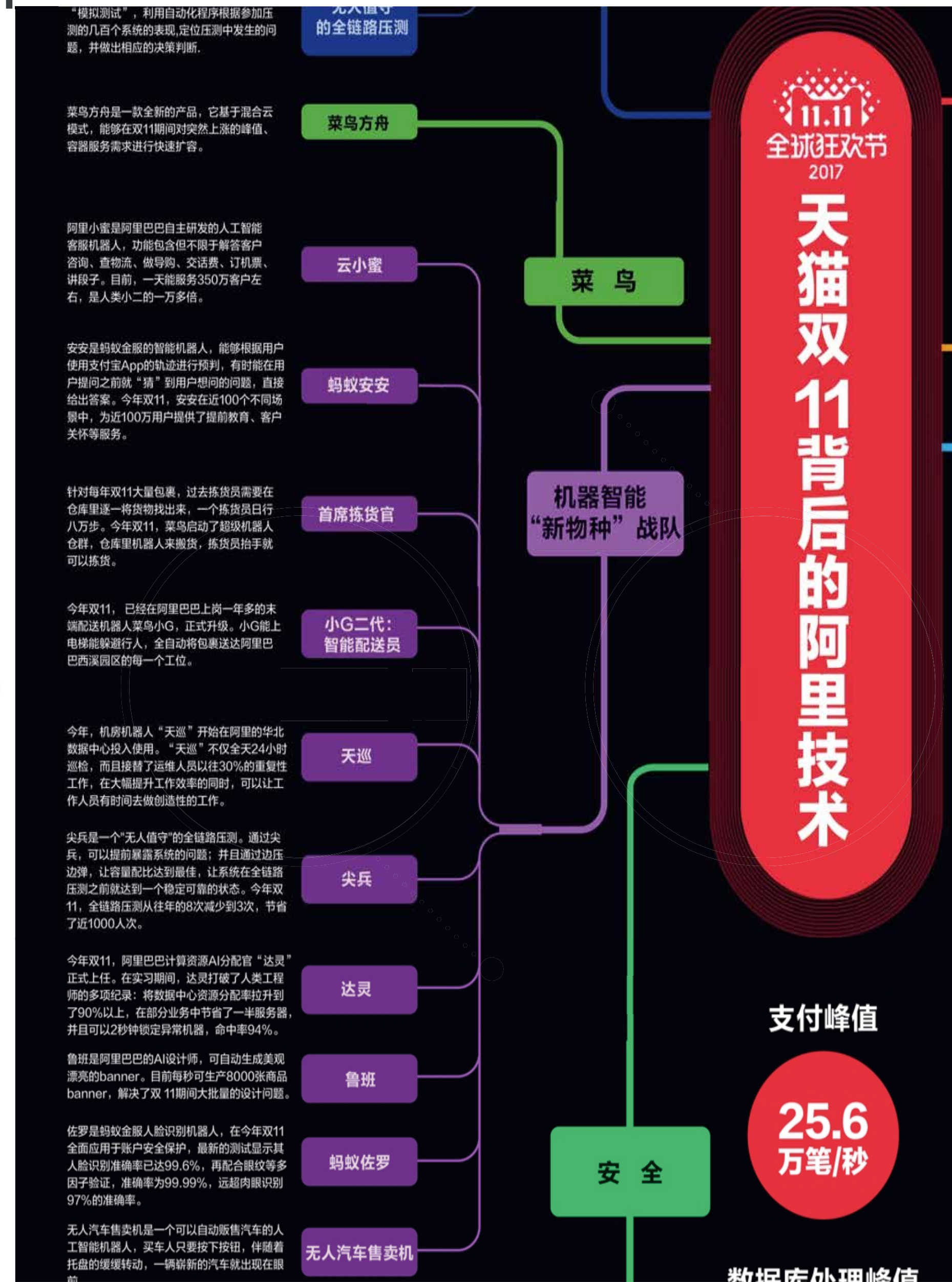
T MALL

理想生活上天猫



最懂商业的技术团队

技术方向：电商技术、智慧供应链、智能营销、机器学习与人工智能应用技术等



机器智能“新物种”战队

小G二代

智能配送员

今年双11，已经在阿里巴巴上岗一年多的**末端配送机器人**菜鸟小G正式升级。小G能上电梯能躲避行人，全自动将包裹送达阿里巴巴西溪园区的每一个工位。升级版的小G更智能更敏捷，很快就会走出园区，给消费者们投递包裹了。



蚂蚁佐罗

——“颜”为定的刷脸智能

佐罗基于**人脸验证核身**，可以给你的账号多加一把锁。它将在今年双11全面应用于账户安全保护，最新的测试显示其人脸识别准确率已达99.6%，再配合眼纹等多因子验证，准确率为99.99%，远超肉眼识别97%的准确率。



达灵

数据中心AI调度官

每年双11都需要保证每一个功能模块都能分配到服务器数量以及监督运行情况。去年双11，就20名工程师负责这样的工作，今年达灵上岗，实现**无人化分配数据资源**，节省了一半的服务器资源。



旗舰舱

首席拣货官

针对每年双11的大量包裹，过去拣货员需要在仓库里逐一将货物找出来，一个拣货员经常要日行八万步。今年双11，菜鸟在仓库内部署了**机器人旗舰舱**，机器人负责将装有商品的货架拉到拣货员跟前，拣货员抬手就可以拣货，一天最多走三两千步。



尖兵

单挑压测特种兵

为了保障剁手党们流畅的**买买买**，在双11之前尖兵会化身成为模拟流量洪峰的特种机器人，挑战系统承压能力，它可自己对容量配比进行智能调整，快速定位问题，得益于尖兵，今年天猫双11的全链路压测节省了1000人次的工程师。



鲁班

机器人设计师

在双11前，鲁班已经学习数百万设计图像，可以根据运营提供的产品元素**智能生成Banner**，每秒可生产8000张商品Banner，解决了双11期间大批量的设计问题。



店小蜜

智能客服

阿里为商家开发的一键授权的客服机器人，一年365天*24小时**不间断工作**。从双11预售开始，店小蜜单日对话量都在1000万以上，是首个日对话量达到千万量级的服务机器人，刷新历史纪录。过去，一个店小二每天服务200-250个客户，现在店小蜜一天能服务350万客户左右。



天猫智选

机器运营小二

天猫智选是一套**智能货品选品运营系统**。它通过大数据分析出什么样的商品天生就带着“爆款”基因，能够成为双11最受欢迎的尖货。这名“机器小二”能够像真正的运营小二一样，指导卖家科学备货，甚至通过与其他系统的配合，为热门商品带来更多的流量。



机器导购员

智能推荐系统

基于对你的了解，智能推荐系统发成为了懂得每个人喜好的“金牌导购员”，能够实现“**千人千面**”，为你推荐最符合你要求的商品。智能推荐系统今年将进一步进化，变得更加聪明，不仅能猜测出你最想要的商品，在搜索关键词的时候，它也能迅速的为你推荐最符合你品味的商品。



天巡

机房巡逻员

全天24小时巡查数据中心，排查异常，能对**亿万级**的数据点进行秒级检测和故障定位，接替了运维人员以往30%的重复性工作，目前已在**一机房实现无人作业**，未来可以独立看管2万台服务器、规模达2万平方米以上数据中心。



双11会场



选品/选商家

如何从10亿的商品中，挑选出合适的商品，从百万卖家中，挑选出最优质的商家

销量预测

需要机器学习技术做销量预测，指导商家备货、指导天猫进行全局运营，把更好的商品投放给消费者

会场构建

需要图像技术、自然语言处理技术，在TB级别文本数据&图像中挖掘场景&主题，智能化构建会场

营销利益点生成

需要通过深度学习技术，自动生成营销导购内容&素材

个性化推荐

需要个性化推荐算法，根据消费者全网行为大数据，构建千人千面的个性化导购会场

深度学习

自然语言
处理

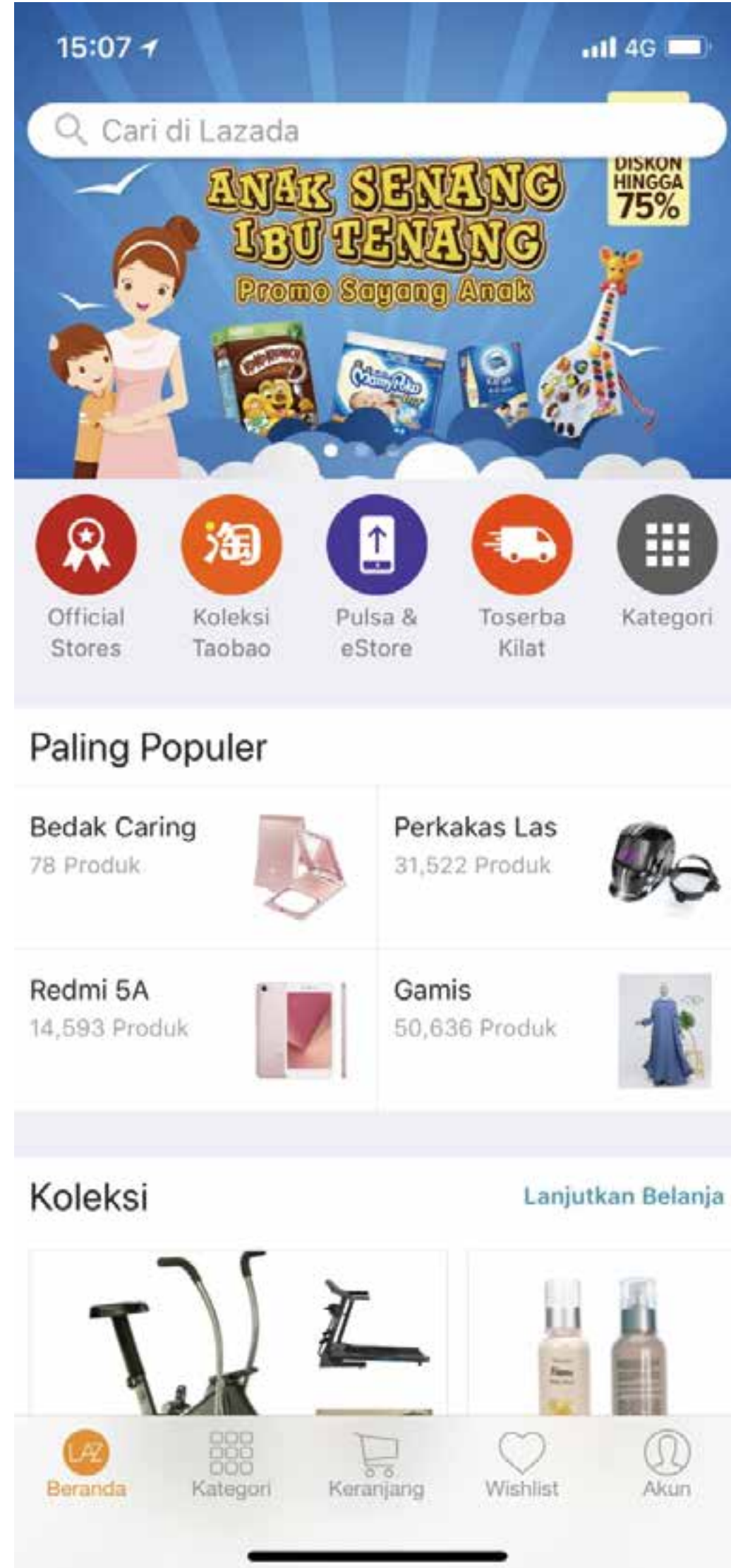
计算机视觉

销量预测

智能会场

个性化推荐

海外市场



需求发现/知识挖掘

基于海外大量的消费者行为数据进行需求发现与预测,针对搜索和社会化数据利用NLP能力进行知识挖掘和发现

用户画像

针对数亿商品、6亿海外消费者进行画像

智能定价

针对商品进行针对性的智能定价已满足市场竞争要求

品类规划

针对全网10亿商品进行品类规划



需求挖掘

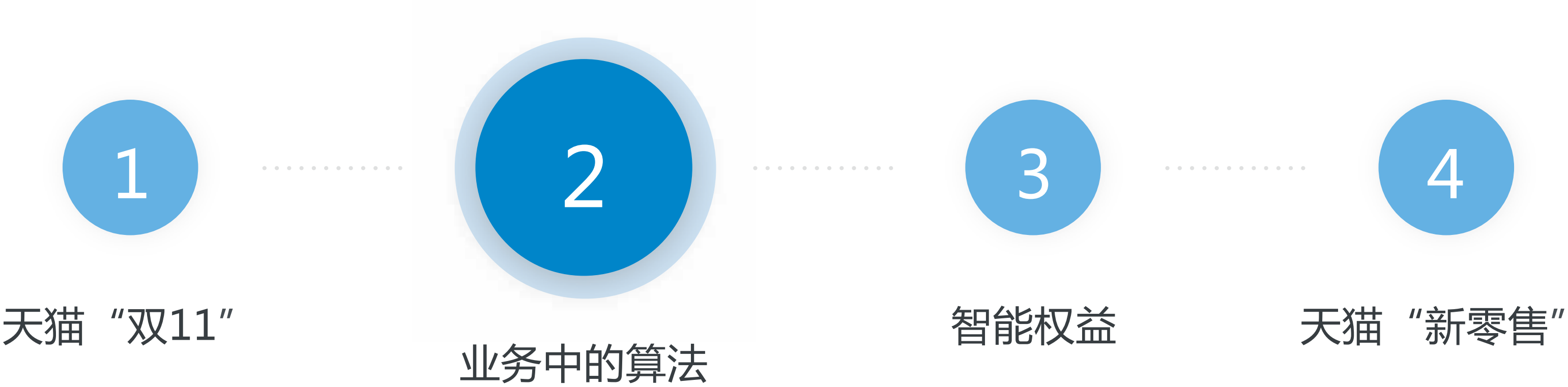
知识挖掘

用户画像

智能定价

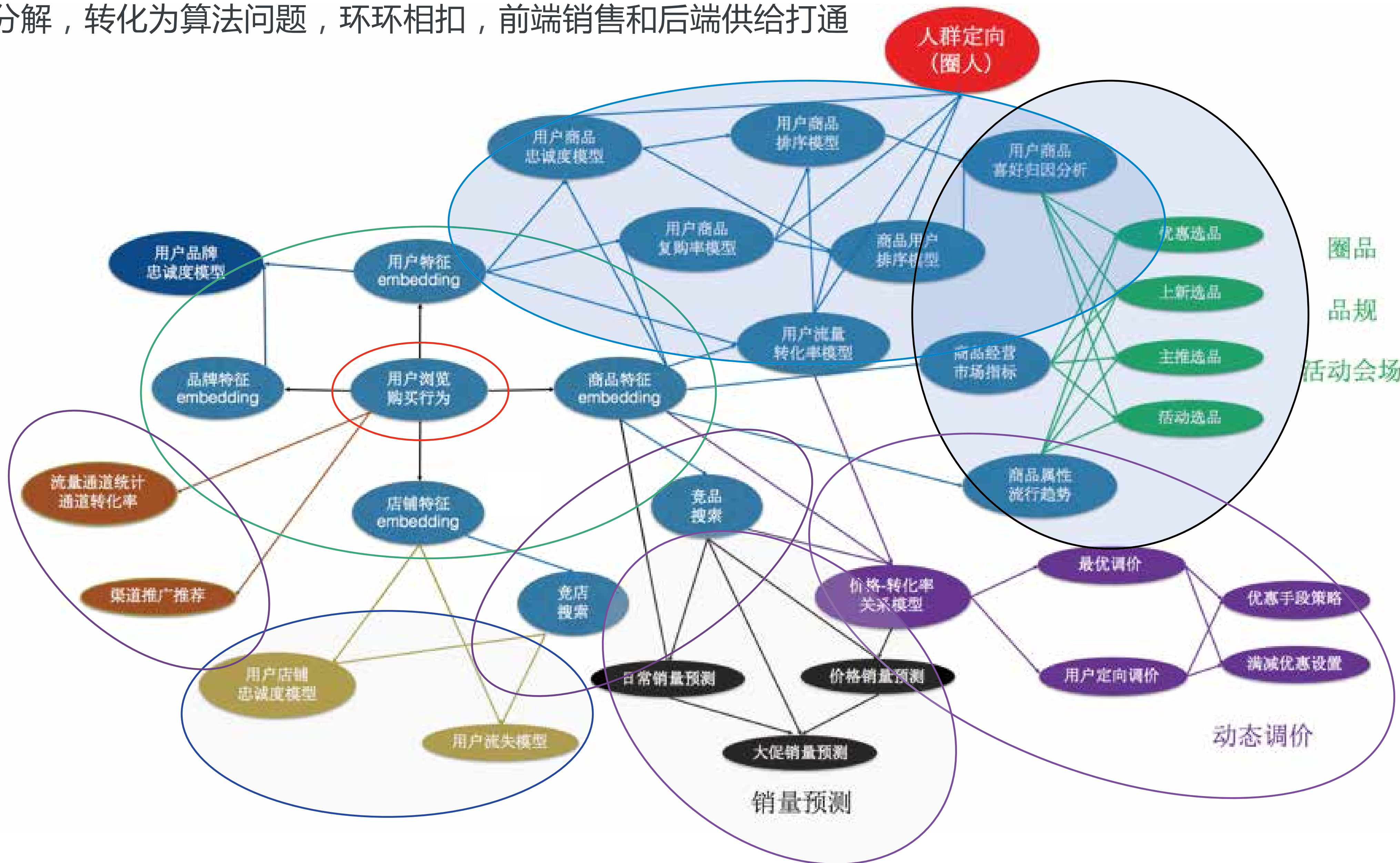
品类规划

目录



算法模块关系图

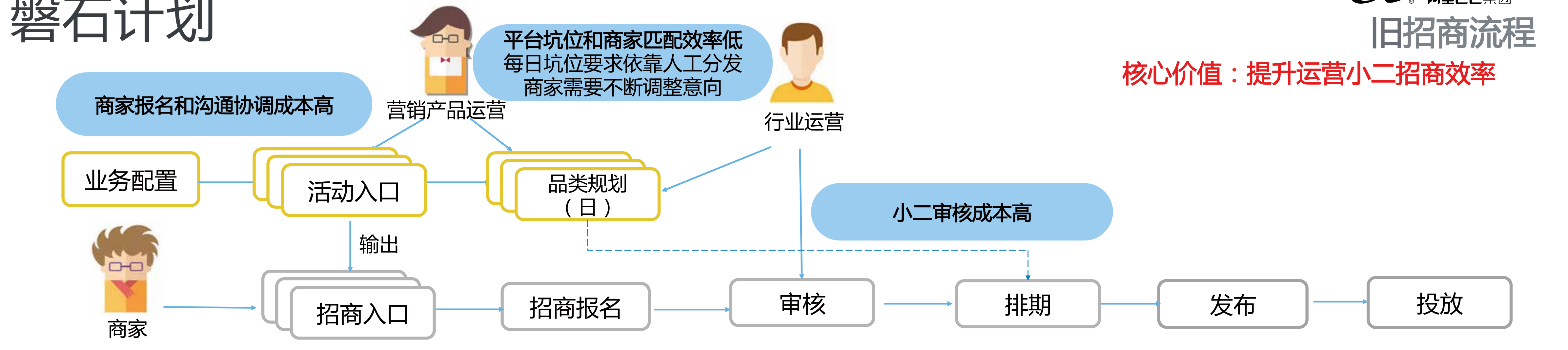
业务问题分解，转化为算法问题，环环相扣，前端销售和后端供给打通



磐石计划

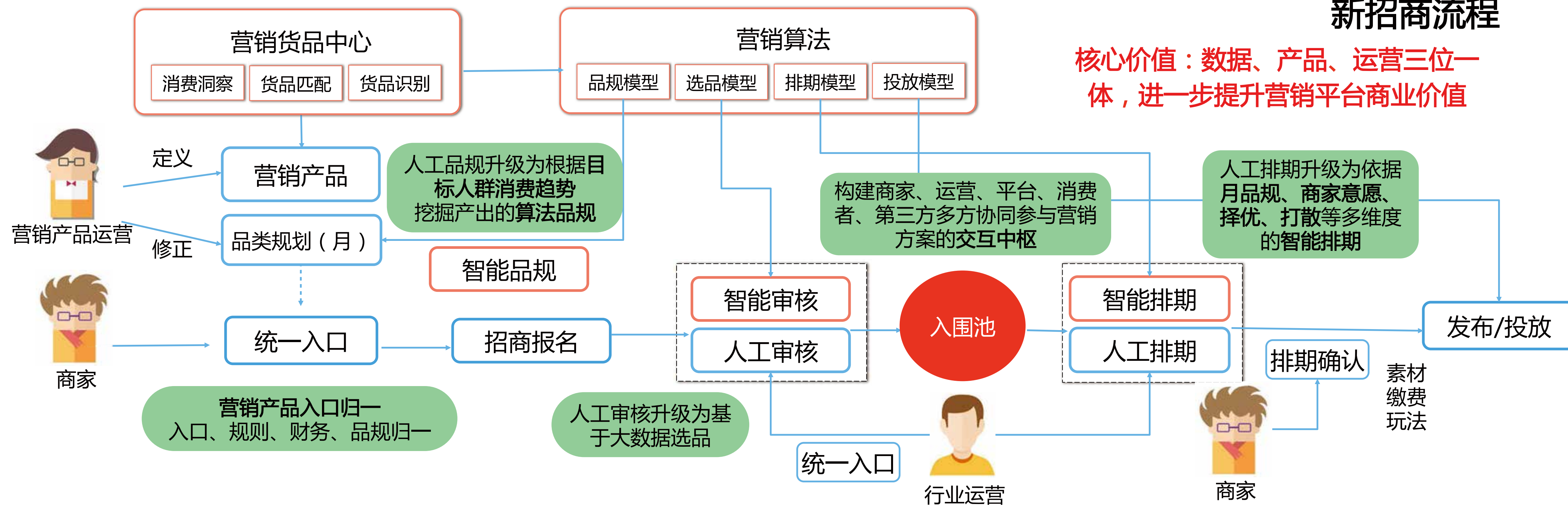
旧招商流程

核心价值：提升运营小二招商效率



新招商流程

核心价值：数据、产品、运营三位一体，进一步提升营销平台商业价值

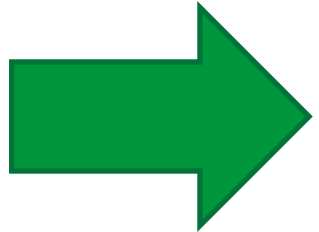


品类规划算法

品类规划-业务需求因子

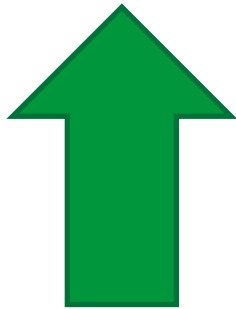
- 1) 坑位单产
- 2) 营销产品目标人群
- 3) 大盘类目成交趋势
- 4) 类目颗粒度

叶子类目坑位

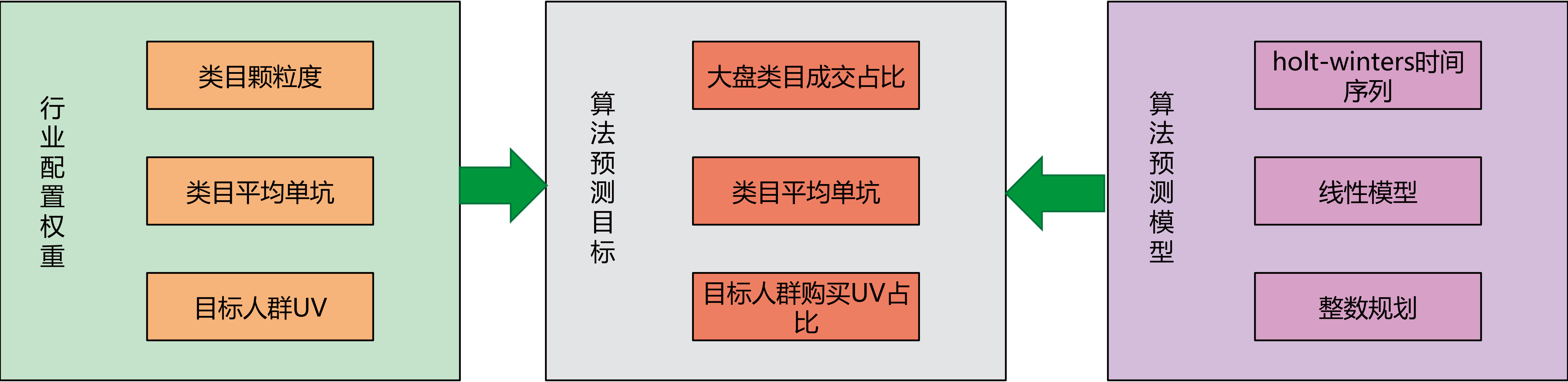


向上汇总

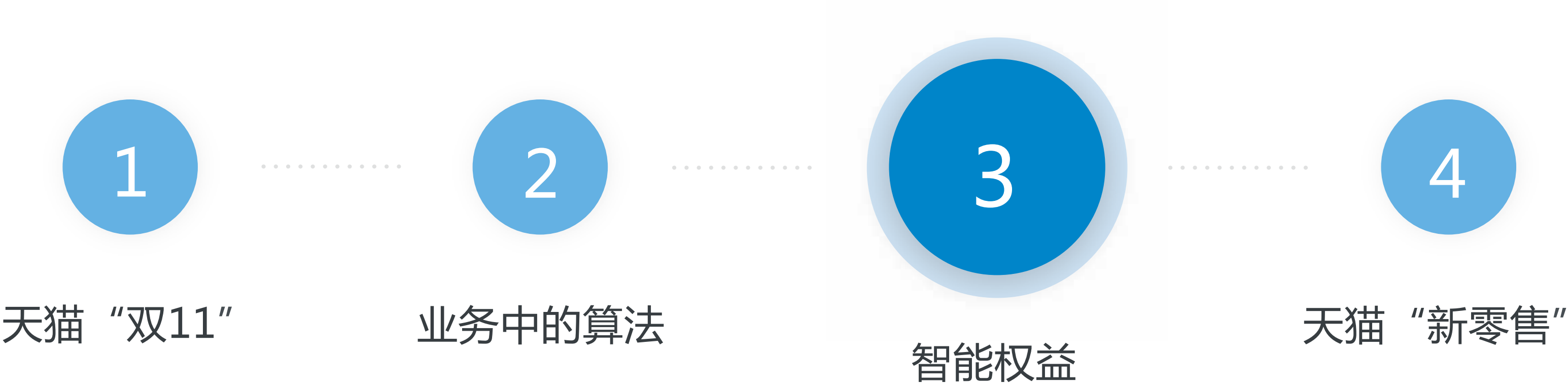
类目	坑位数
女装	10
手机	5
连衣裙	6



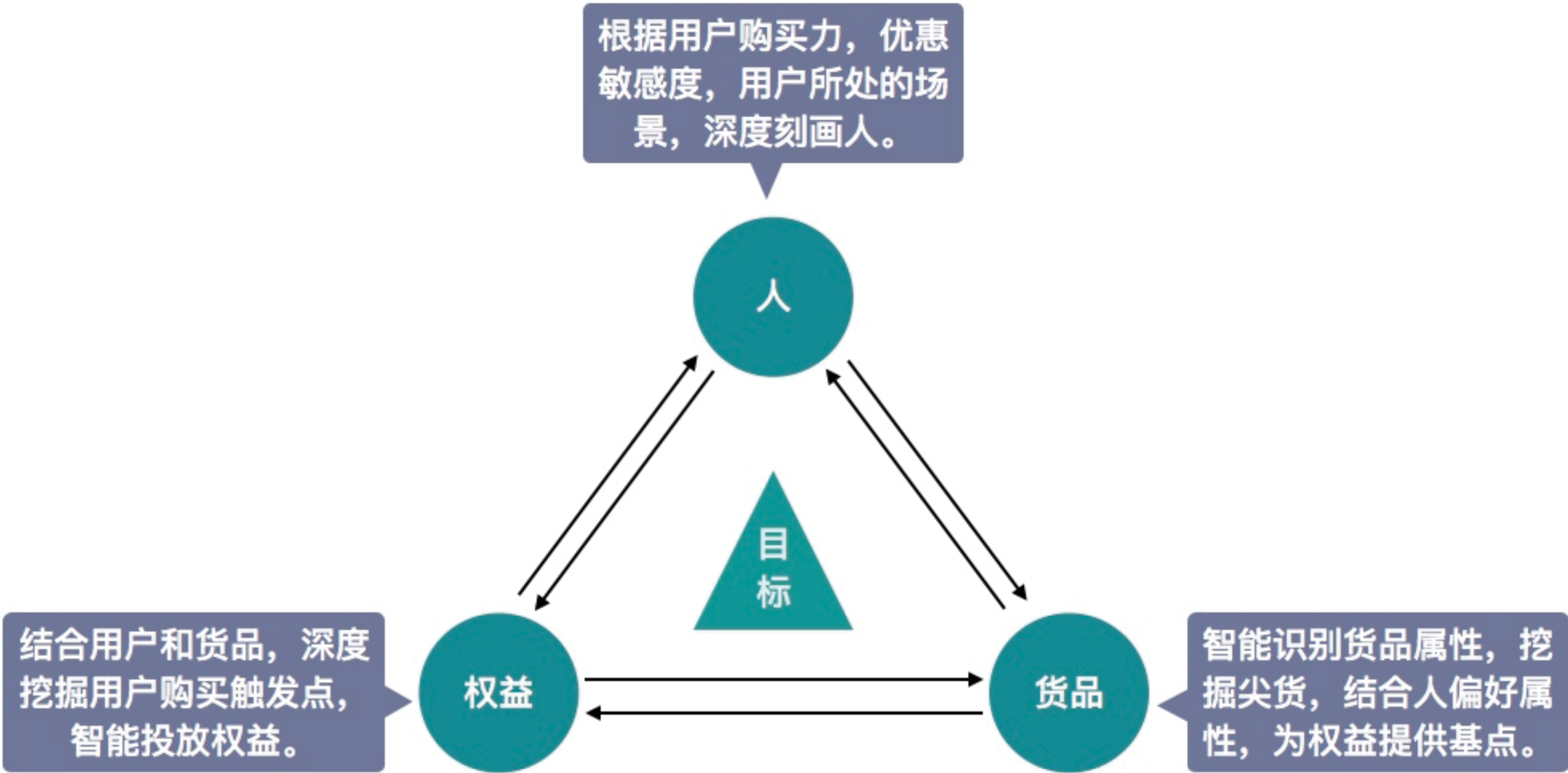
输出



目录



权益智能发放



三位一体打造全面的智能化场权益投放，基于对权益&人&货品三个维度认知基础上，挖掘两两间更深度的关系，三位一体构建智能化权益发放。

双11智能权益



基础模型

双11当天消费金额预估模型

- 消费者属性特征
- 消费者行为特征
- 历史购买金额

消费者优惠敏感度模型

- 消费者特征
- 历史购买记录
- 优惠金额

消费者传播力模型

- 消费者历史互动行为
- 用户传播力因子
- 用户活跃度因子

消费者类目偏好模型

- 用户特征
- 类目特征
- 用户&类目特征

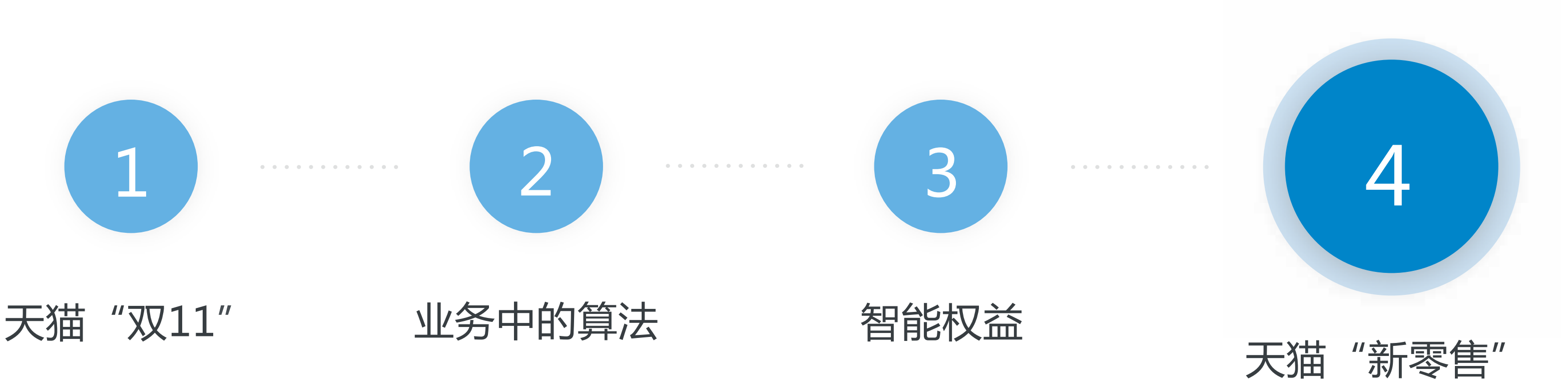
流失用户模型

- 平台基础特征
- 用户个人属性特征
- 用户的行为数据

发放策略

针对不同的营销目的，组合使用多种模型，形成相应权益类型的发放策略

目录



图形图像算法

- 服饰/家装等设计属性识别：什么属性流行，变化是什么，消费者偏好是什么，为下一期商品企划做决策参考

算法模型：定义美（细化到设计师的语言），发现美（主色 / 廓形 / 关键点 / 设计元素 / 设计风格识别），创造美（搭配推荐，风格合成，虚拟试衣）。并且，结合统计销量 / 搜索量 / 浏览量等需求趋势，预测各设计元素的流行趋势，以指导设计

- 智能终端（试衣镜、试妆镜等）：包括人脸关键点检测，体型检测，视觉呈现优化等。
- 智能终端/智慧门店：
 1. 基于视觉方案的客流统计及顾客动线收集
 2. 用户行为识别及分析
 3. 基于视觉方案的货品数量监控及补货提醒

“无人零售店” 淘咖啡



天猫智慧快闪店



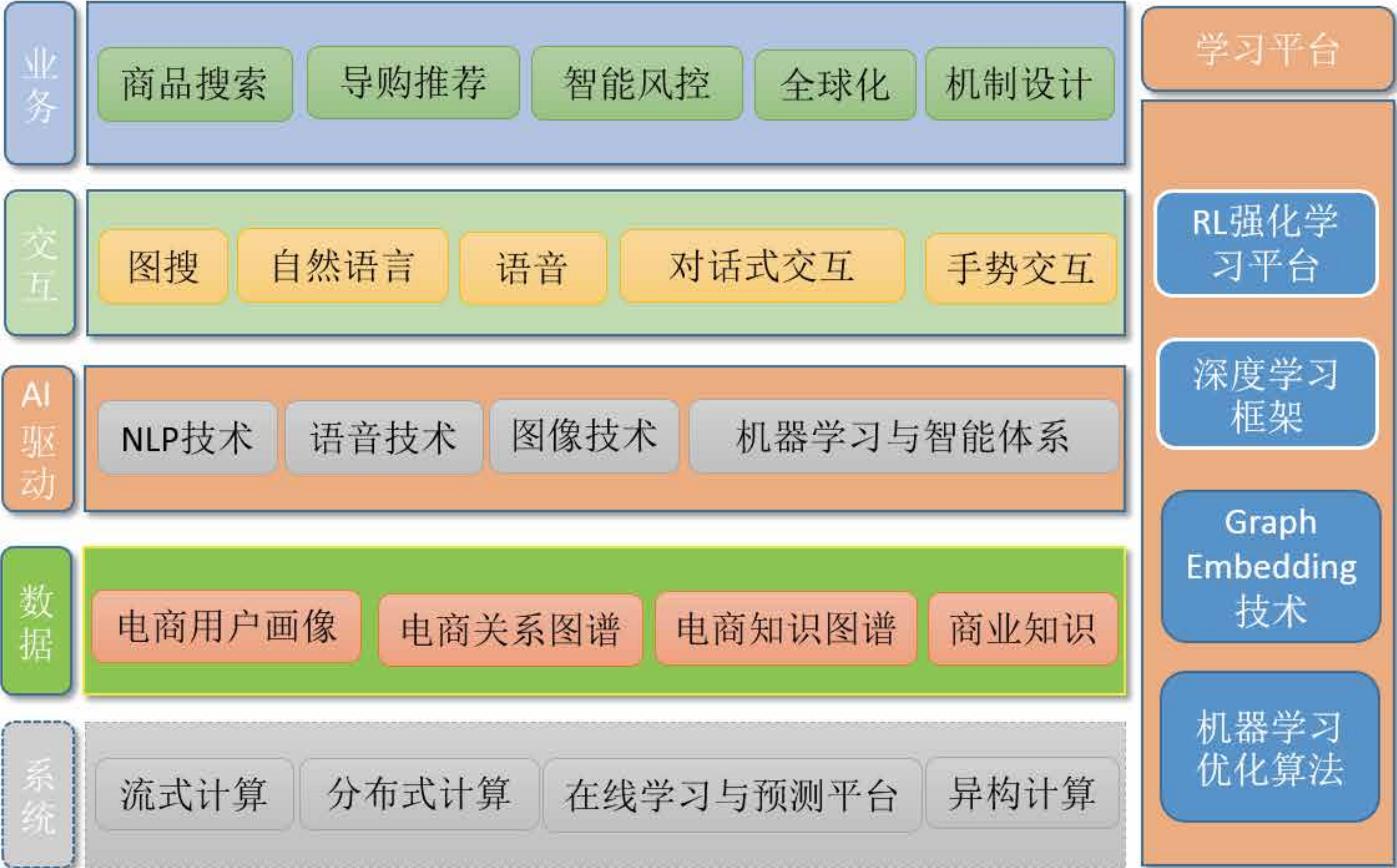
搜索事业部：AI@搜索和推荐

资深算法专家 三桐

特点

- 输入
 - 最大规模的电商数据
 - 数据闭环，商品，浏览，点击，成交，评价，物流全链路数据
 - 数据高度结构化
- 目标
 - 平台，商家多方共赢的最佳用户购物体验
- 方法
 - 算法：实时个性化智能化
 - 计算与工程：实时流式学习和计算

技术图



最早应用于大规模
商业产品的
AI技术

最完整的静态与动态
**电商用户
兴趣图谱**

新
智能交互

最大的电商
关系图谱

流量最大的
**搜索和推
荐服务**

1 最早应用于大规模商业产品的AI

MIT科技评论评选的2016年度10大突破科技中提到在强化学习上，阿里巴巴和其它公司等不同, 阿里巴巴率先在搜索等商用领域取得突破



View from the Marketplace

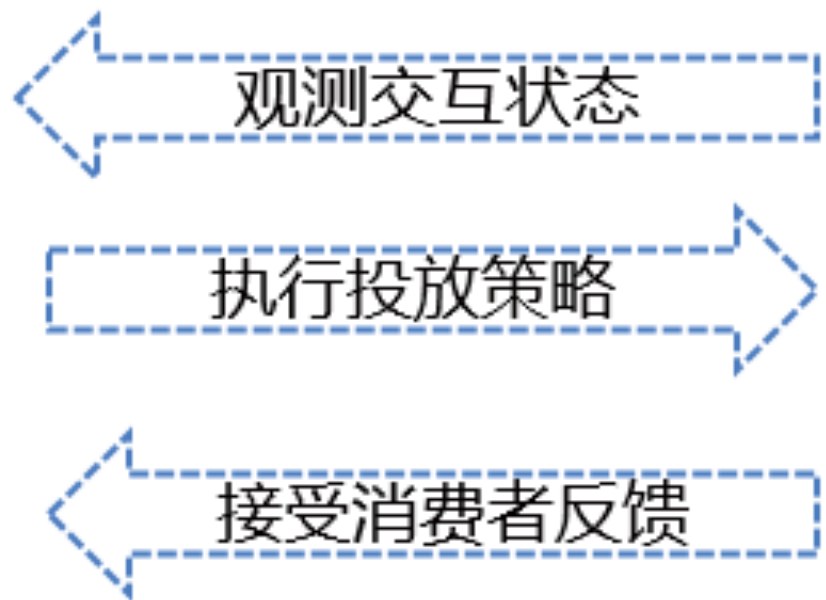
Big Data Game-Changer: Alibaba's Double 11 Event Raises the Bar for Online Sales

Overall, these complex technologies work together to facilitate shopping by providing more choices and greater ease of ordering. In e-commerce search alone, applying deep reinforcement learning and online learning would increase GMV by more than 10 percent on Double 11 day.

1 搜索决策与强化学习 (reinforcement learning)

搜索智能决策：探索 and 收集样本，根据环境反馈学习最优策略

搜索引擎



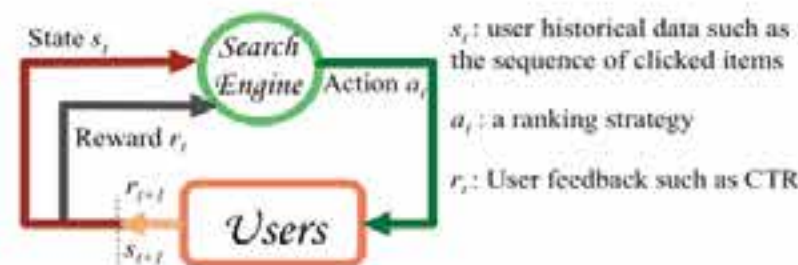
满意
或
不满意

探索和利用
商品冷启动，个性化体验和流量均衡等

Multi-armed
Bandits

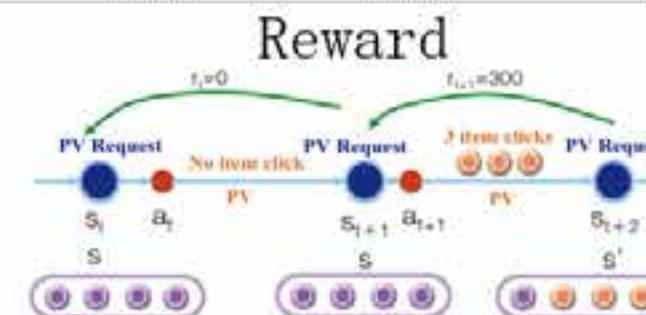
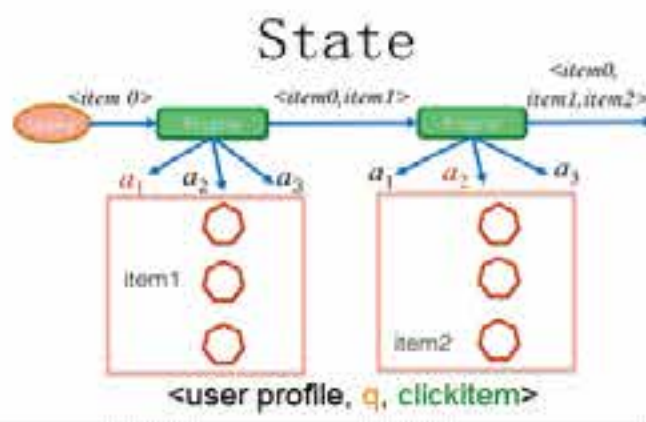
长短期收益平衡
超参数优化
关联场景联合优化等

Deep
Reinforcement
Learning



MDP: $\langle S, A, R, S' \rangle$

$$Q(s_t, a_t) = Q(s_t, a_t) + \gamma(r_{t+1} + Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))$$



Action

排序参数学习

$$score = (w_1 * f_1 + w_2 * f_2 + w_3 * f_3 + \dots)$$

($s = W^T F$)

超参数学习

$$score = (w_1 * v_1 * f_1 + w_2 * v_2 * f_2 + w_3 * v_3 * f_3 + \dots)$$

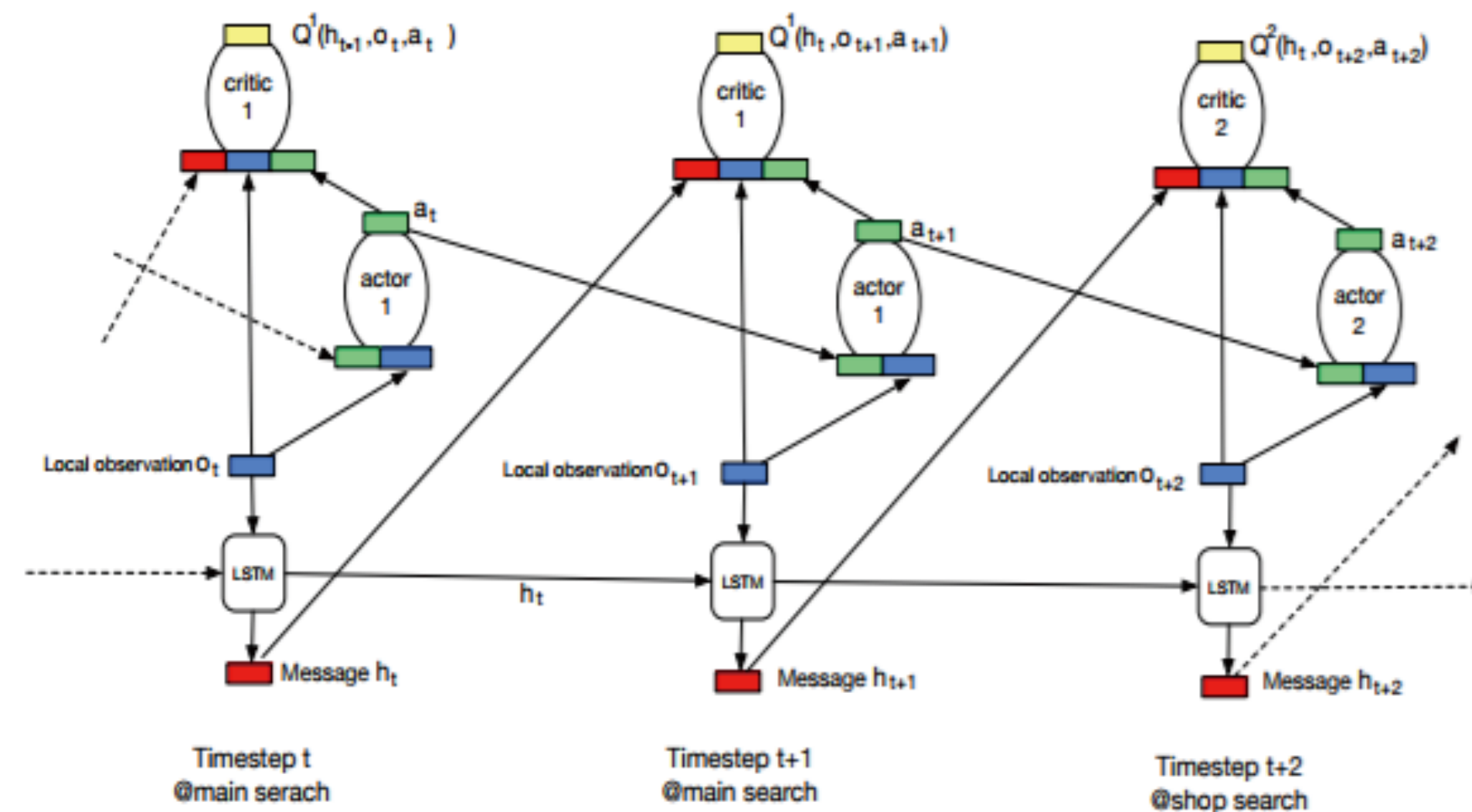
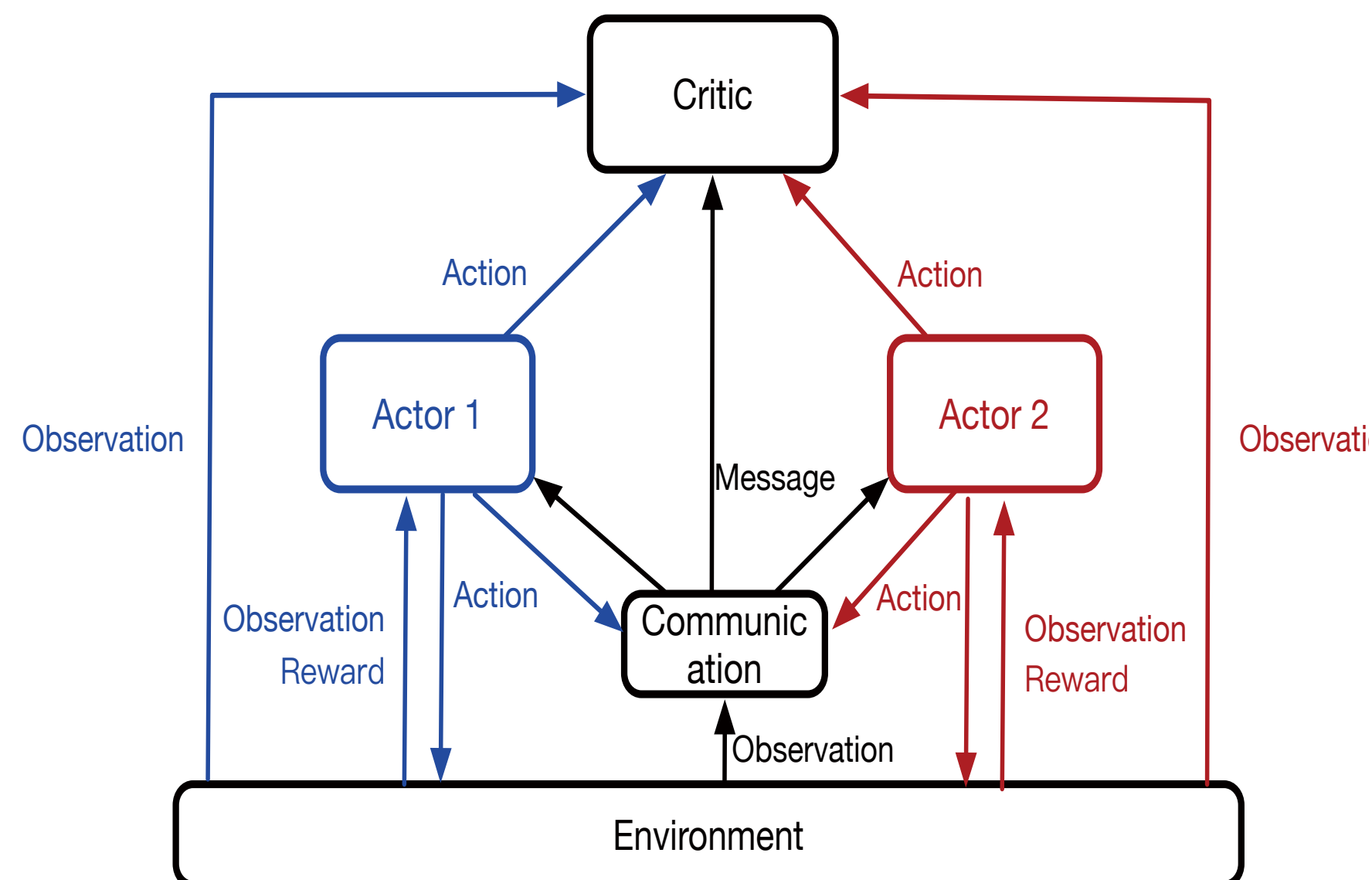
($s = W^T \text{diag}(V) F$) v 取值为0/1

$$score = price^{-t} * (w_1 * f_1 + w_2 * f_2 + w_3 * f_3 + \dots)$$



1 multi-agent learning

- Multi-agent reinforcement learning: 产品A与产品B作为独立两个agents
- 多agent的状态和action统一编码进行智能体间通信
- agent之间状态相互影响，统一目标学习各自的决策策略



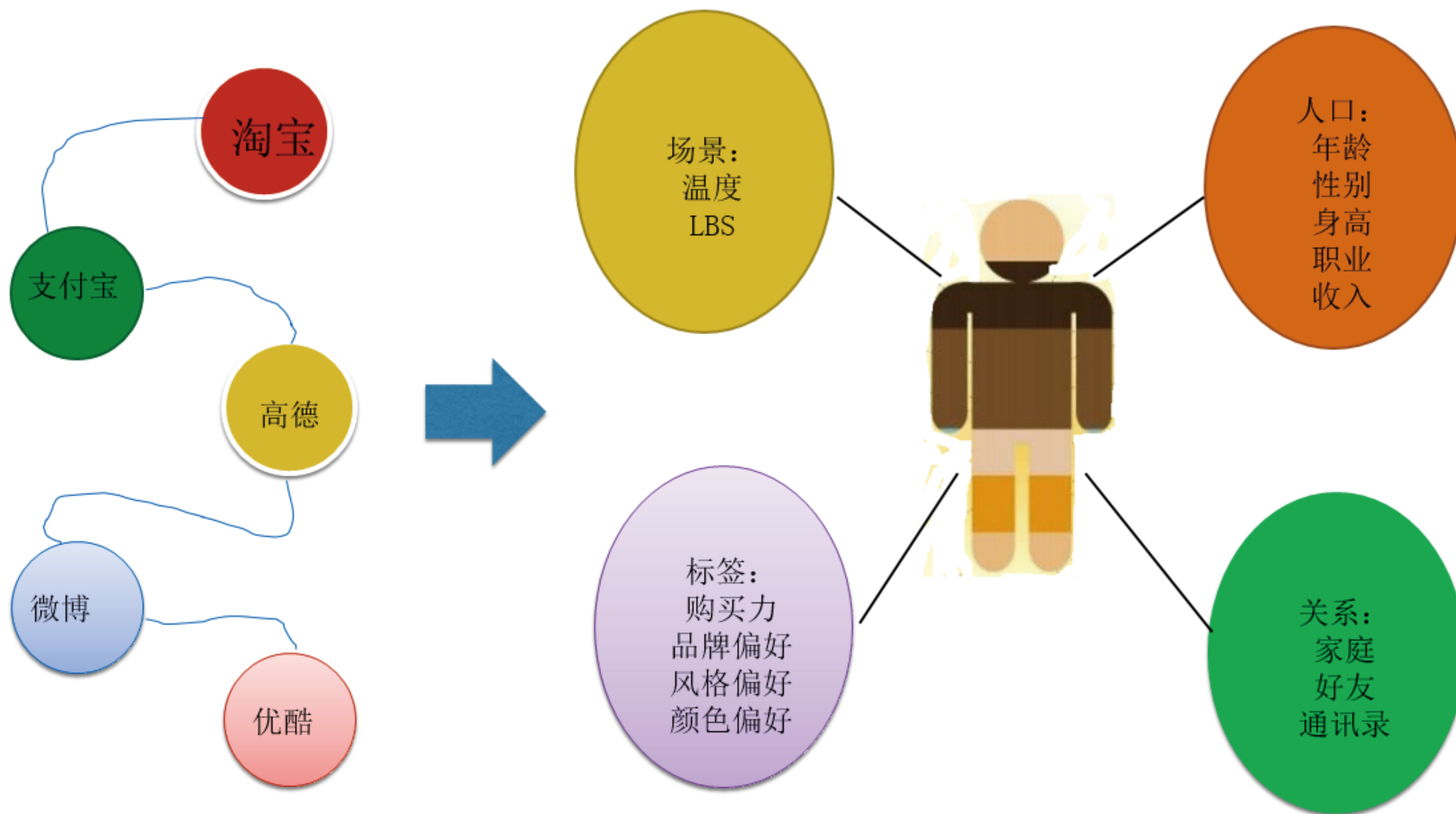
合作学习网络

多actor, 场景 独立决策

集中critic, actor 之间的协同

LSTM, 场景之间通信

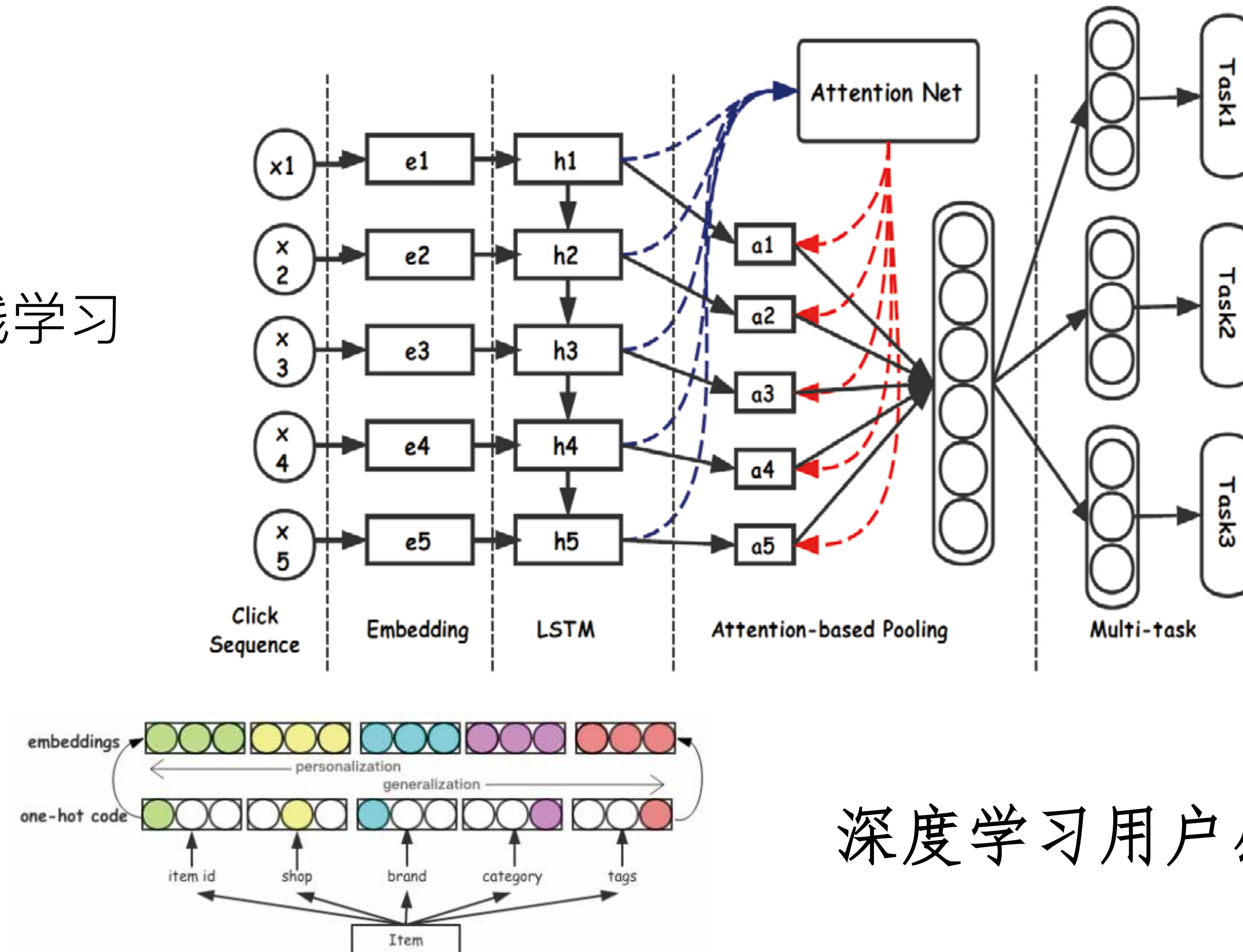
2 最完整的静态与动态用户购物兴趣图谱



2 最完整的静态与动态用户购物兴趣图谱

深度用户感知网络- Deep User Perception Network

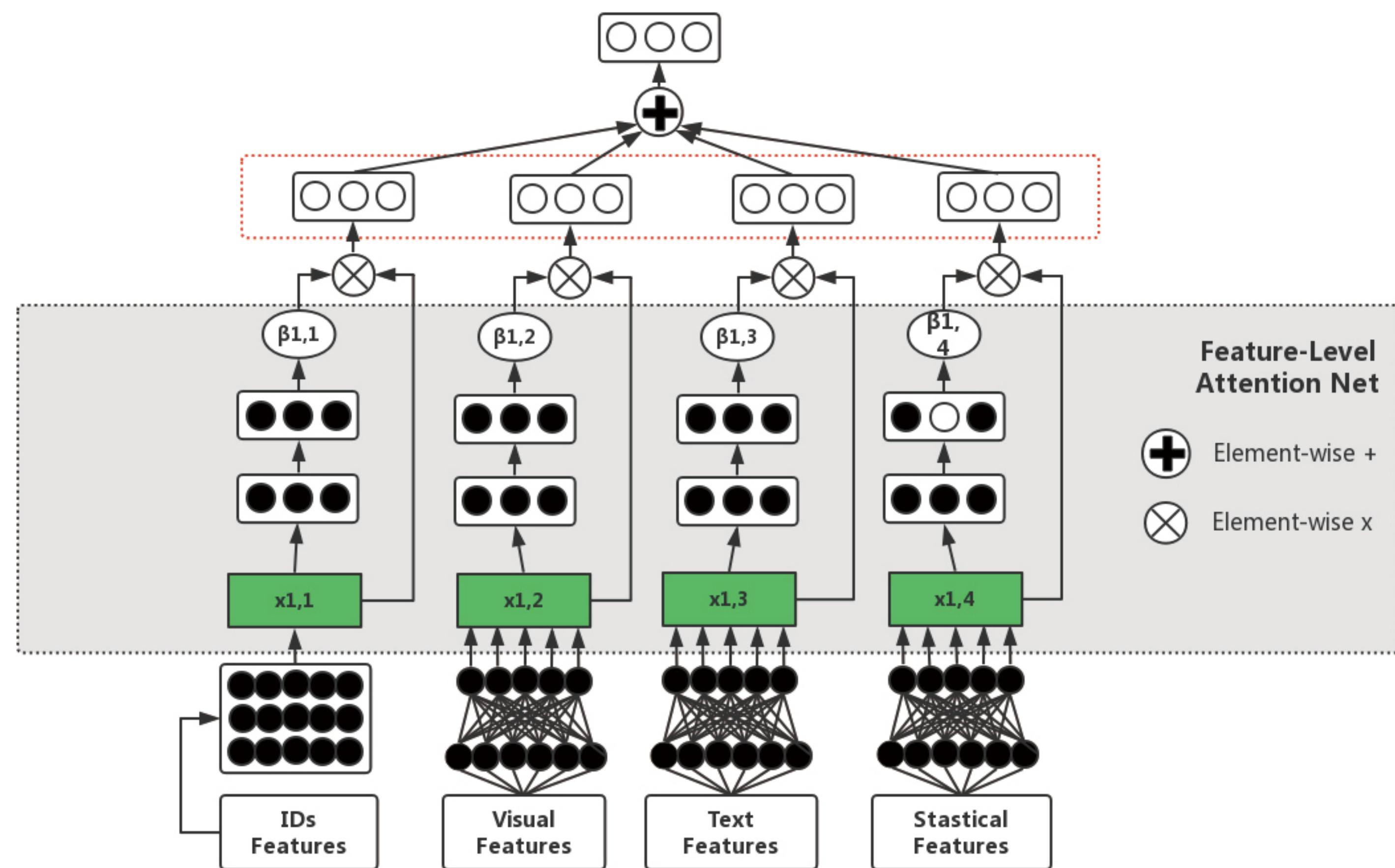
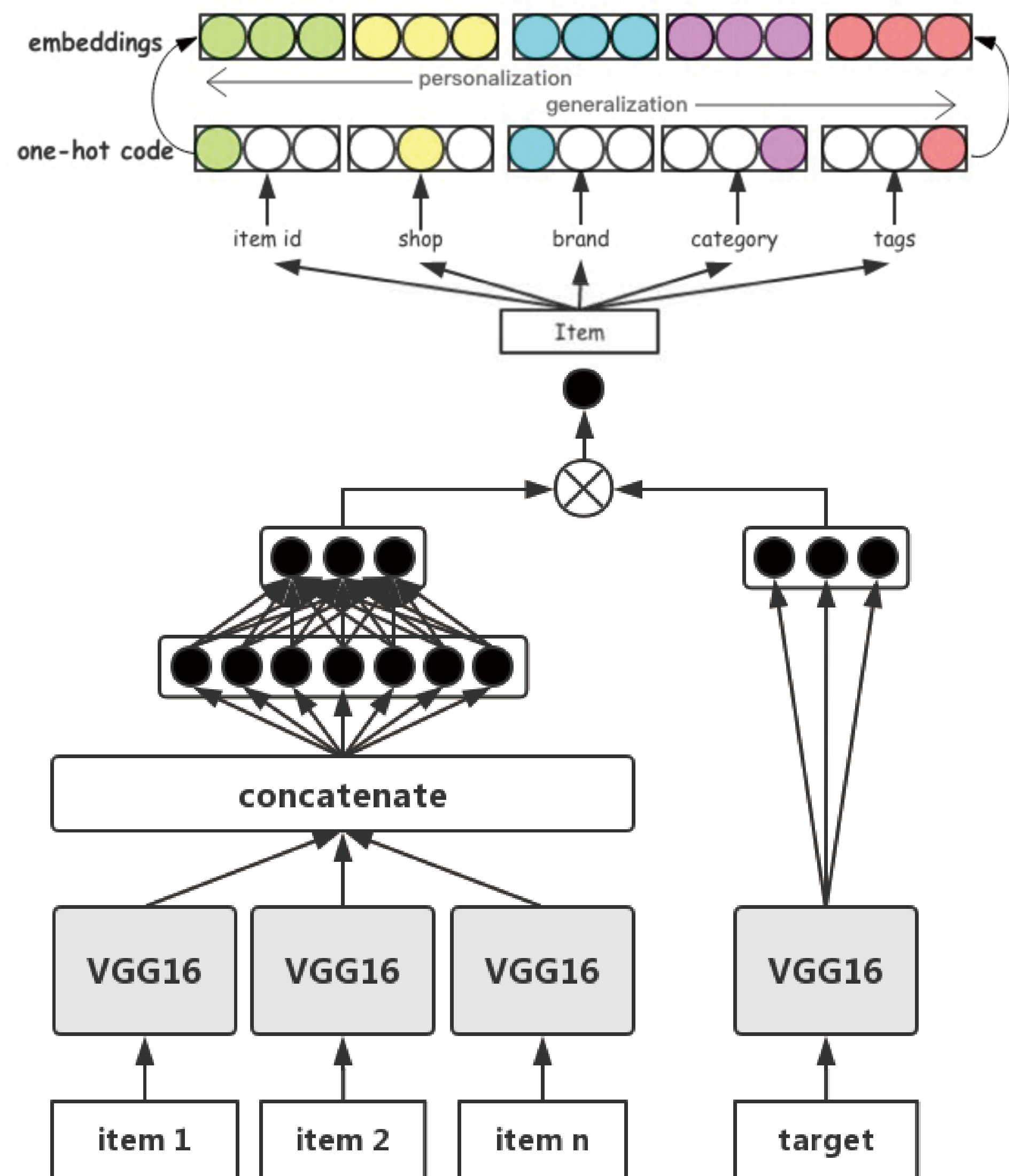
- 基于实时行为序列
- 上下文相关注意力
- 容易迁移到新任务
- 多任务统一表征
- 10B 参数，双11在线学习



深度学习用户感知网络

2 最完整的静态与动态用户购物兴趣图

深度用户感知网络- Deep User Perception Network



3 新智能交互探索

文本交互



自然语言理解技术

语音交互



语音技术

图像交互



图像技术

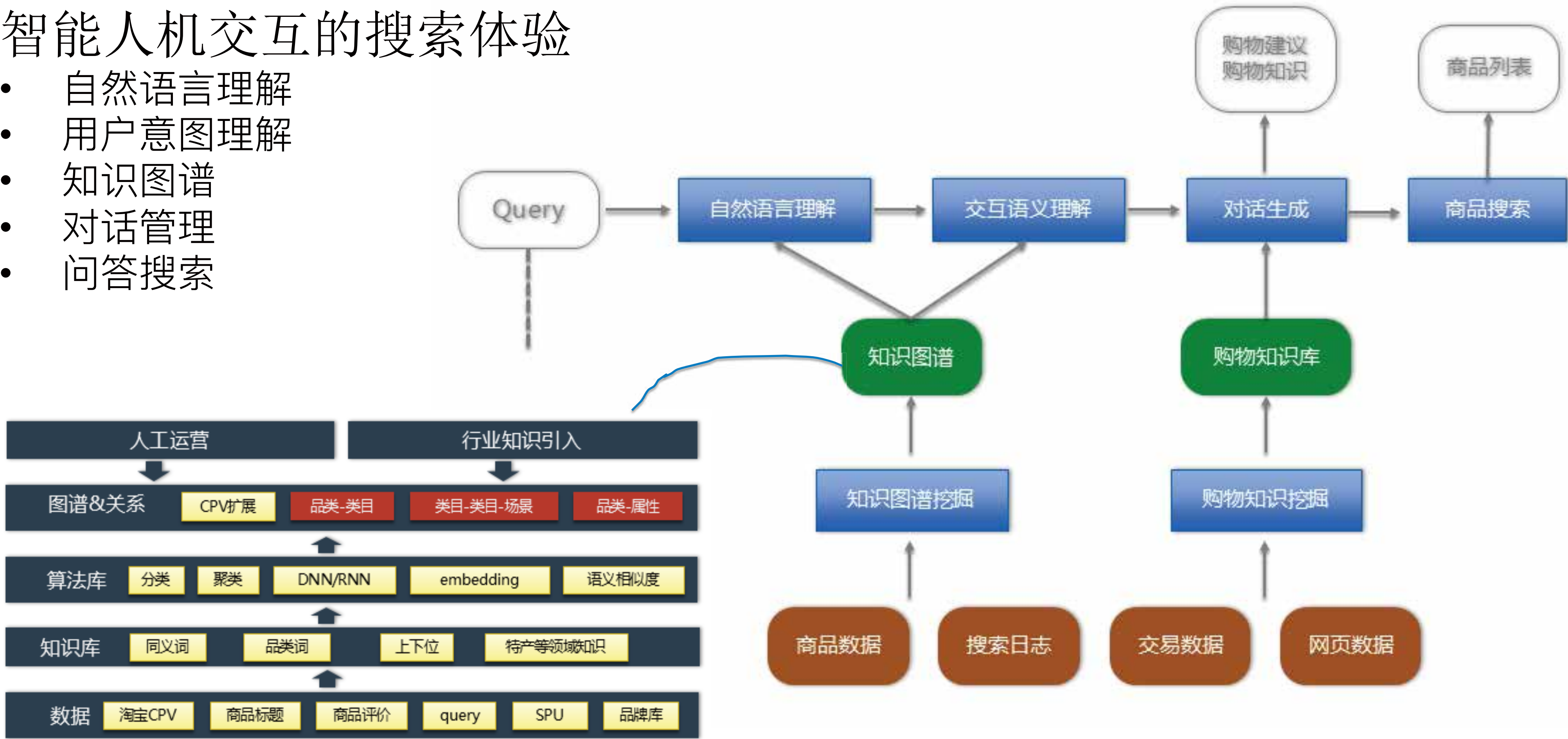
?

自然对话系统?
AR/VR?

3 对话系统

智能人机交互的搜索体验

- 自然语言理解
- 用户意图理解
- 知识图谱
- 对话管理
- 问答搜索



4 最大电商关系图谱

电商图谱，约几十亿实体、千亿边和几十种关系类型，包括

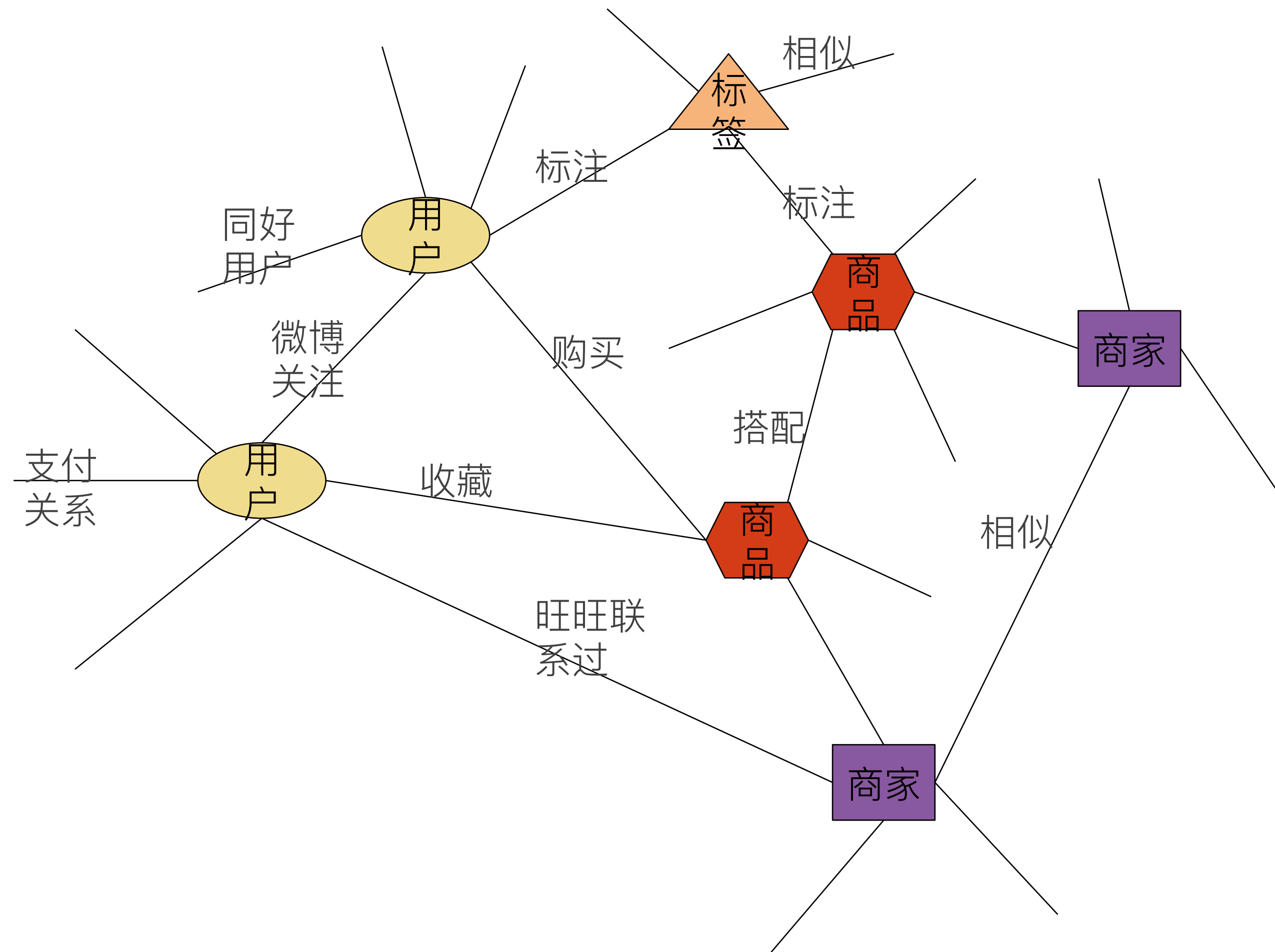
用户社交圈：微博关注、支付、淘宝关注、通讯录、IM通信、家人、好友、同学等

用户生活圈，小区、城市、工作地点、学校、商圈等

用户购物行为，浏览、收藏、加购、购买、评价等

用户标签，偏好类目、偏好品牌、标签等

商品关系，商品相似、同款、搭配、一起点、一起买、凑单；



4 最大电商关系图谱-Deep Graph Embedding算法

应用层

电商关系建设

买卖家分层匹配

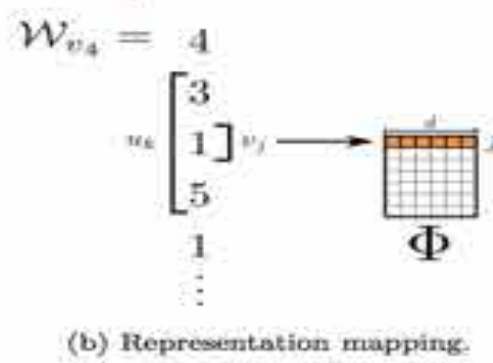
实时反作弊

算法层

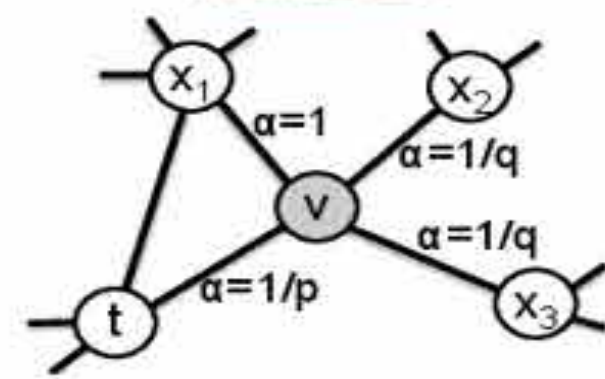
Random-walk-based methods

Deep-learning-based methods

Deep walk

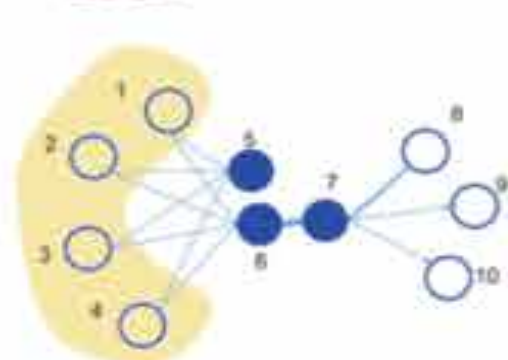


Node2vec



Other methods

LINE



$$\hat{p}_1(i, j) = \frac{w_{ij}}{W}$$

$$p_1(v_i, v_j) = \frac{1}{1 + \exp(-\vec{u}_i^T \cdot \vec{u}_j)}$$

$$O_1 = d(\hat{p}_1(\cdot, \cdot), p_1(\cdot, \cdot))$$

$$O_1 = - \sum_{(i, j) \in E} w_{ij} \log p_1(v_i, v_j)$$

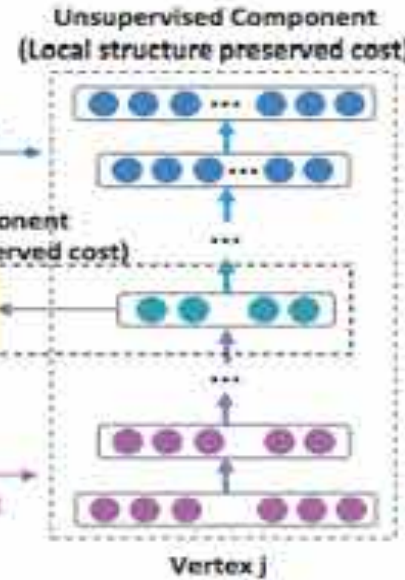
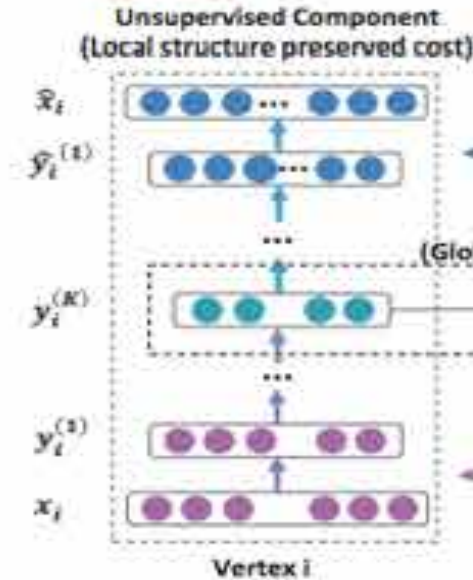
$$p_2(v_j | v_i) = \frac{\exp(\vec{u}_i^T \cdot \vec{u}_j)}{\sum_{k=1}^{|V|} \exp(\vec{u}_i^T \cdot \vec{u}_k)}$$

$$\hat{p}_2(v_j | v_i) = \frac{w_{ij}}{d_i}$$

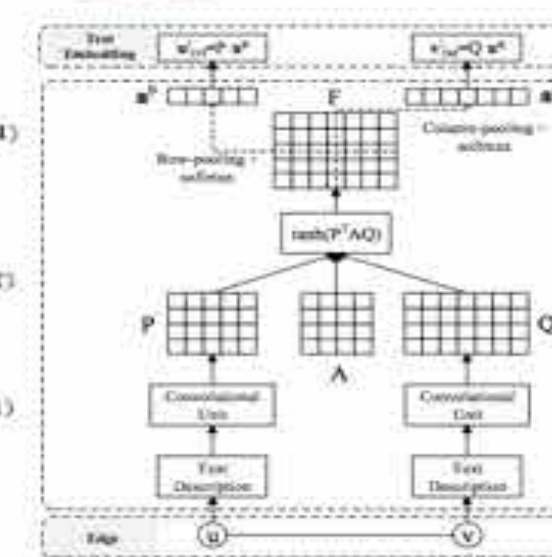
$$O_2 = \sum_{i \in V} \lambda_i d(\hat{p}_2(\cdot | v_i), p_2(\cdot | v_i))$$

$$O_2 = - \sum_{(i, j) \in E} w_{ij} \log p_2(v_j | v_i)$$

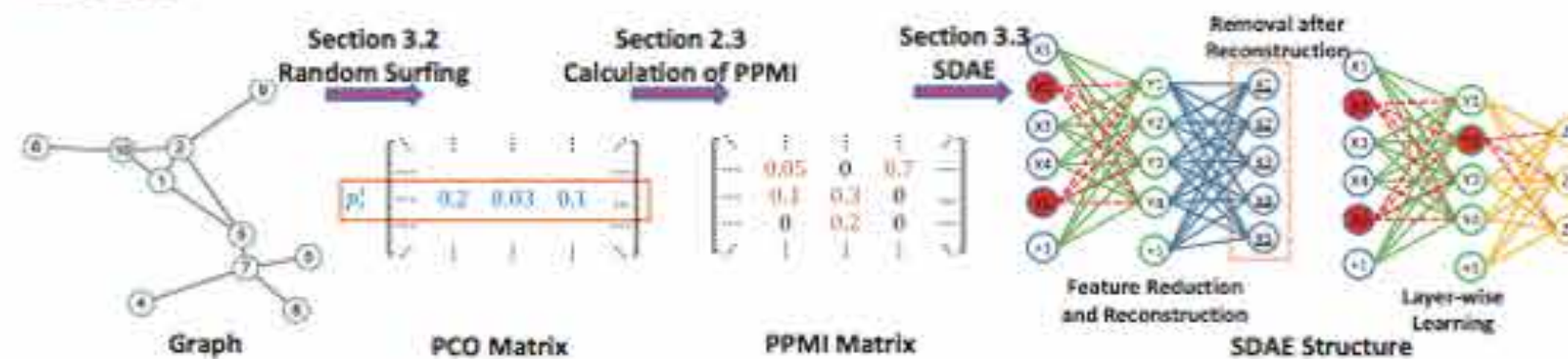
SDNE



CANE



DNGR



系统层

ODPS-GRAPH

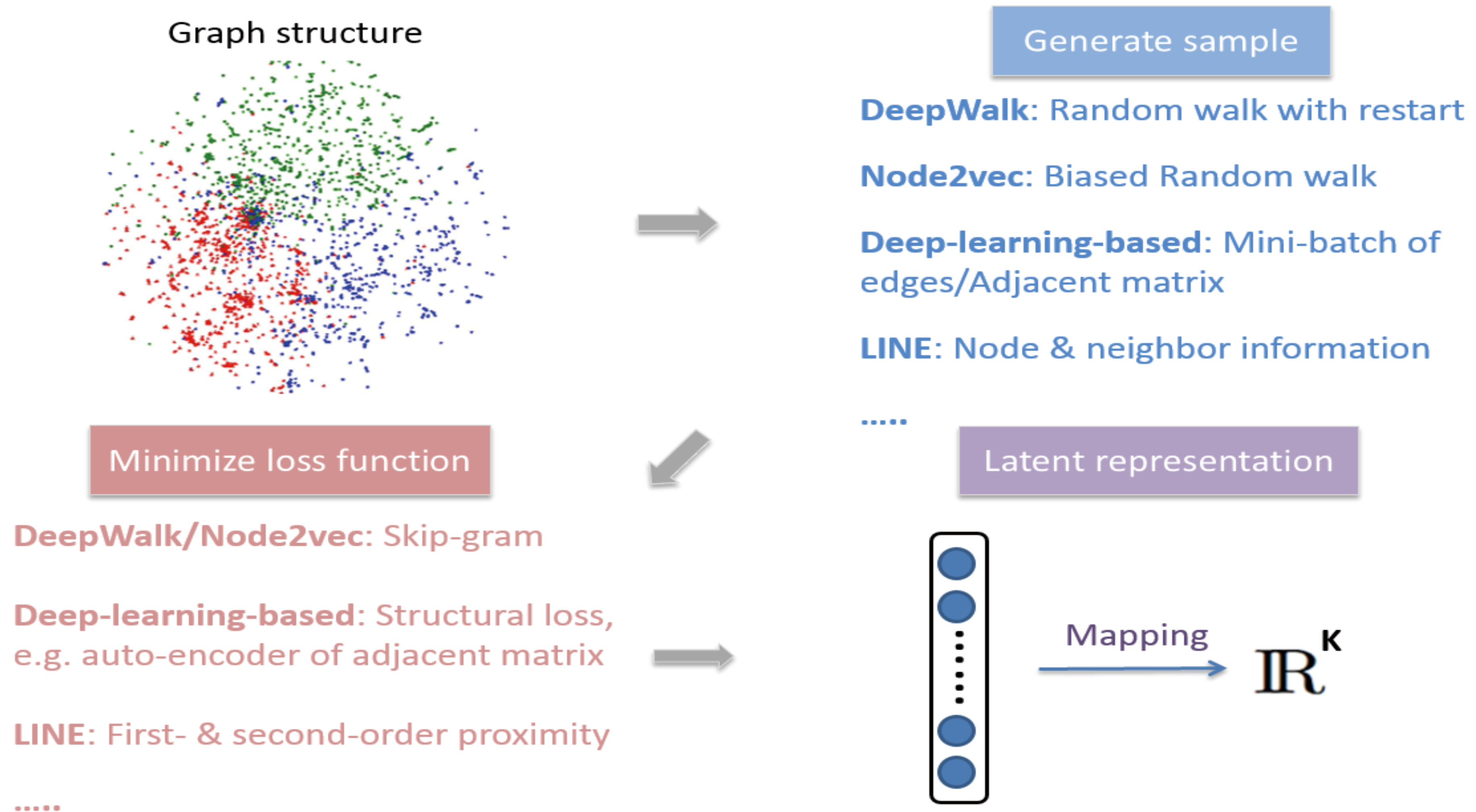
BIG-GRAPH

PAI-TENSORFLOW

PARAMETER SERVER

Embedding
学习节点的隐表示向量
节点关系编码
向量距离表示关系

4 最大电商关系图谱-Deep Graph Embedding算法



5 流量最大的搜索和推荐服务-智能调度

目的:

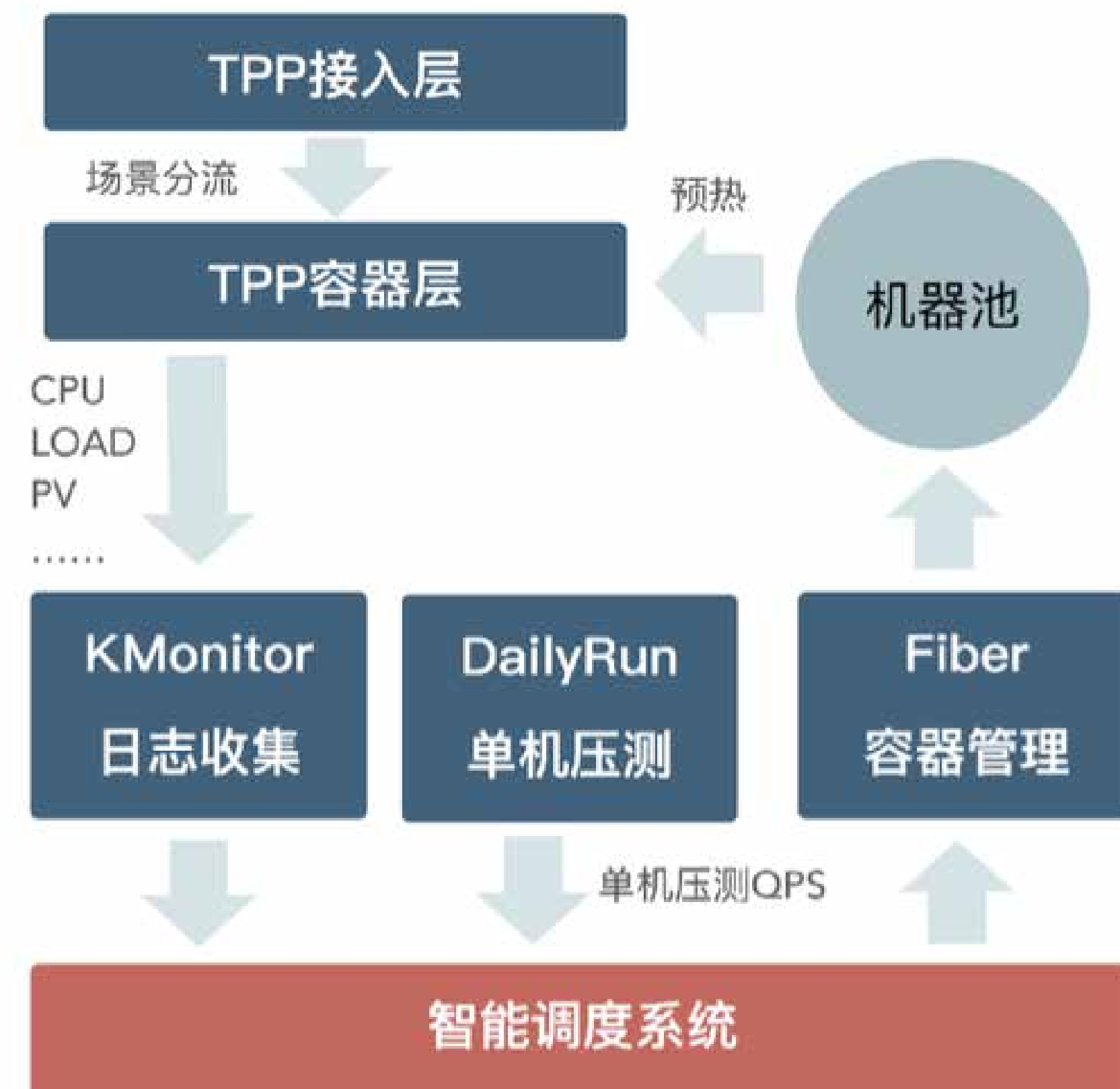
1. 最大化机群使用效率
2. 优先保障重点场景

挑战:

1. 流量突发, 不可预见
2. 扩展和调度时间
3. 全局状态实时感知

方法:

1. 有限资源下的带约束优化问题



机群使用效率提升XX%, 核心场景降级率下降到X%一下

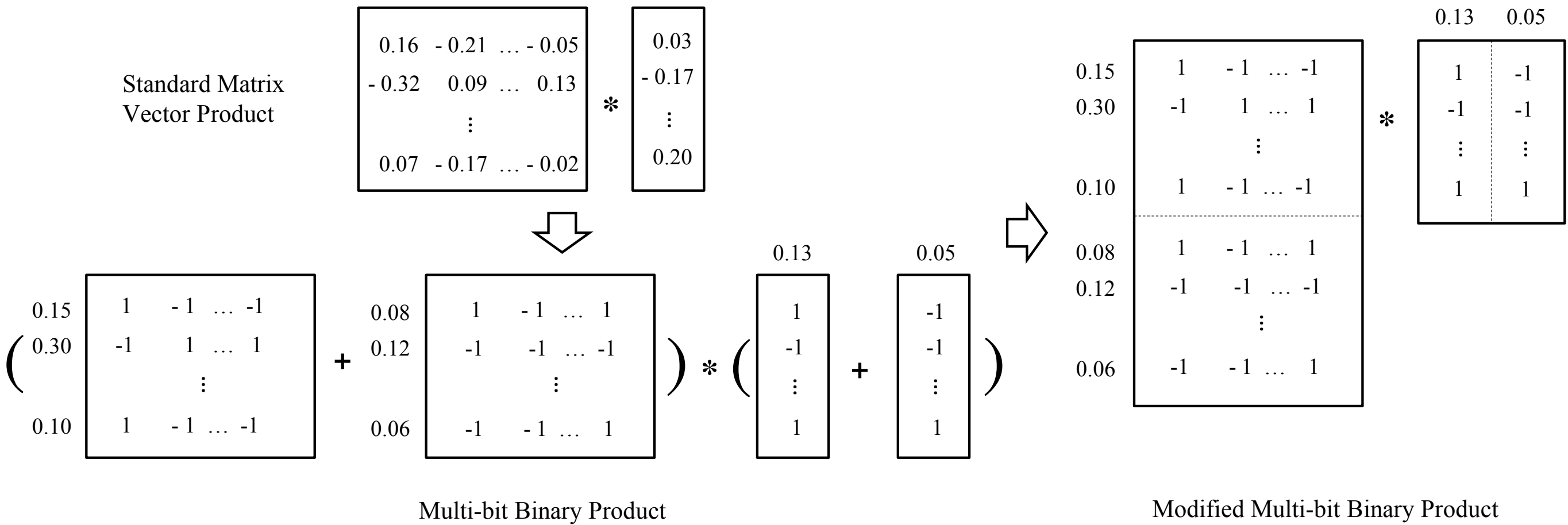
5 流量最大的搜索和推荐服务-加速

加速算法

- 降维
- 量化
- 剪枝
- 二值网络

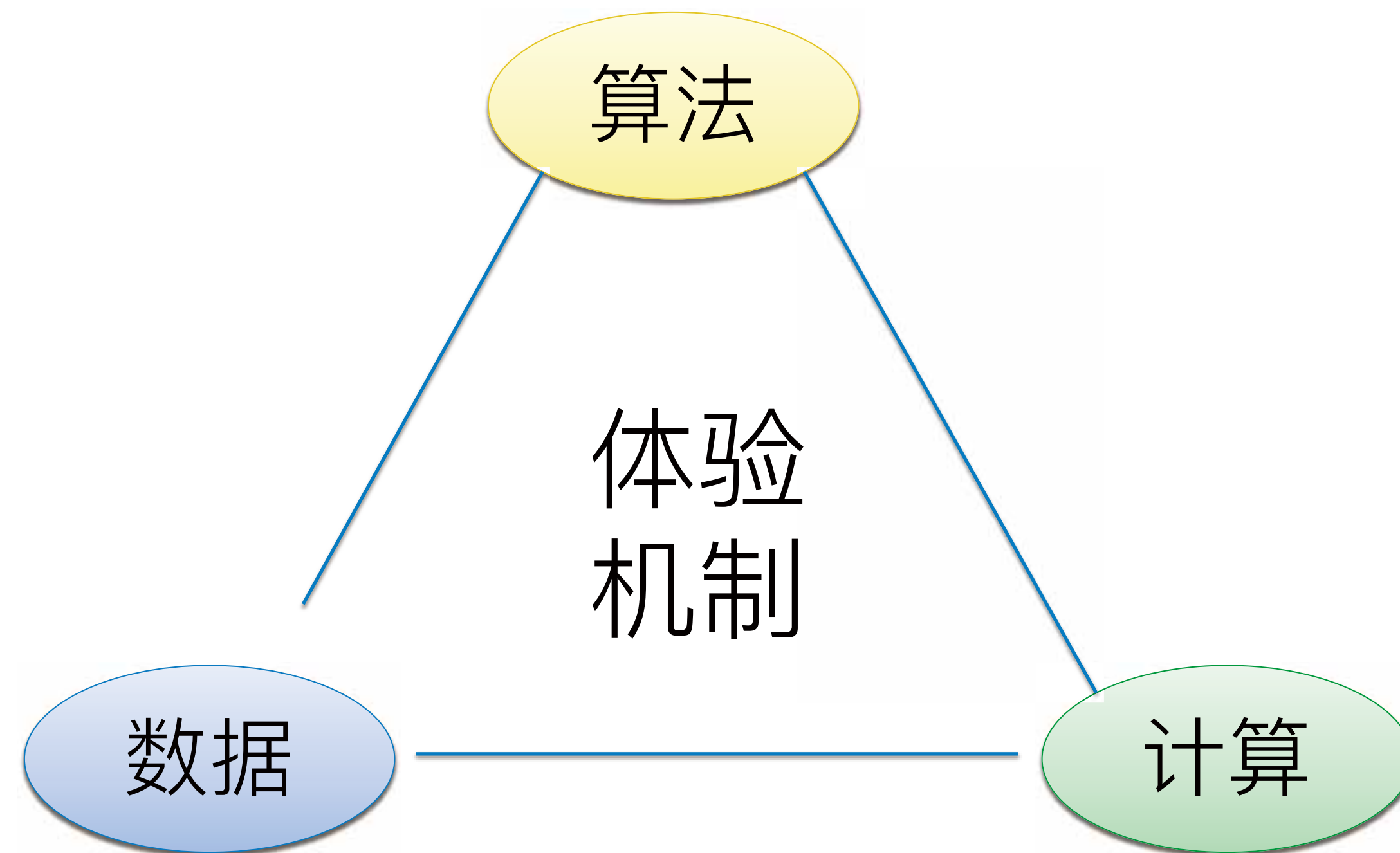
硬件加速

- GPU
- FPGA
- ASIC
- 混布



$$\min_{\mathbf{w}, \{\alpha_i, \mathbf{b}_i\}_{i=1}^k} f\left(\sum_{i=1}^k \alpha_i \mathbf{b}_i\right)$$
$$s.t. \quad \{\alpha_i, \mathbf{b}_i\}_{i=1}^k = \arg \min_{\{\alpha'_i, \mathbf{b}'_i\}_{i=1}^k} \|\mathbf{w} - \sum_{i=1}^k \alpha'_i \mathbf{b}'_i\|^2.$$

挑战



- 研究方向
 - 机器学习，深度学习，强化学习
 - 自然语言理解（NLP），知识图谱
 - 交互：图像，语音，对话式交互
 - 博弈论，安全，机制设计
 - 优化算法
 - 流式计算，图计算，异构计算，混布计算

新零售供应链平台事业部： 智慧供应链

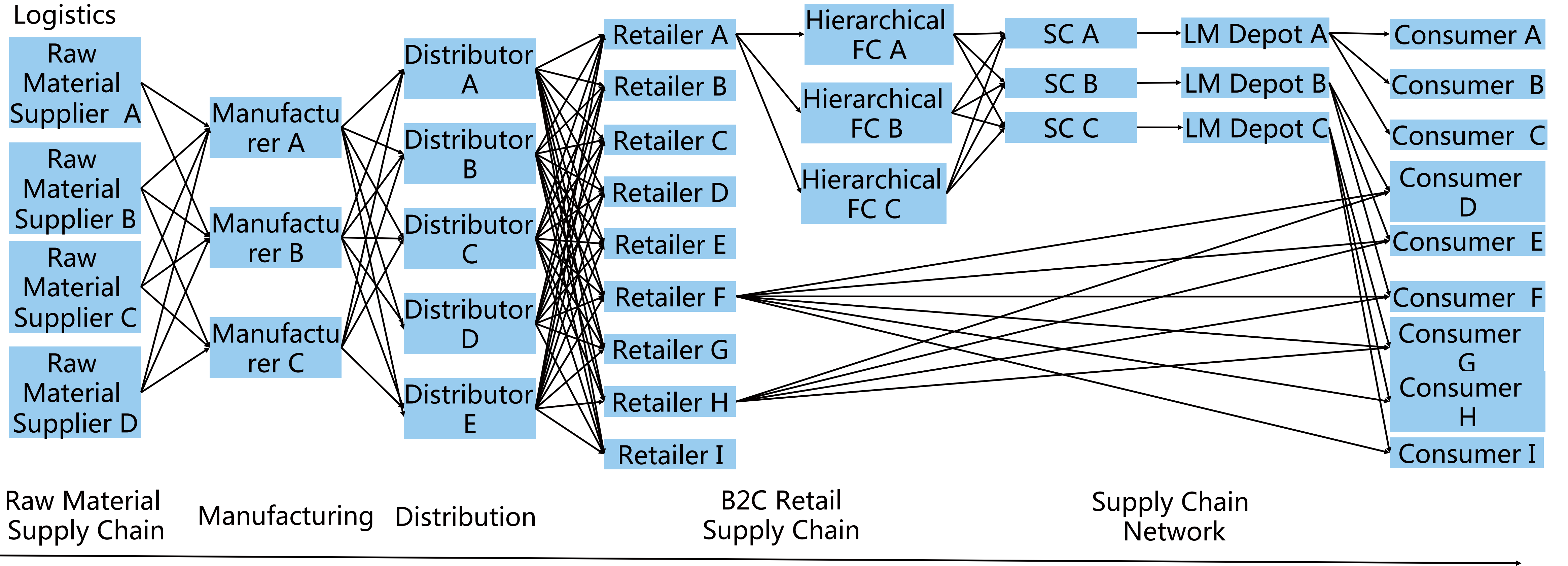
资深技术专家 粤谦

目录



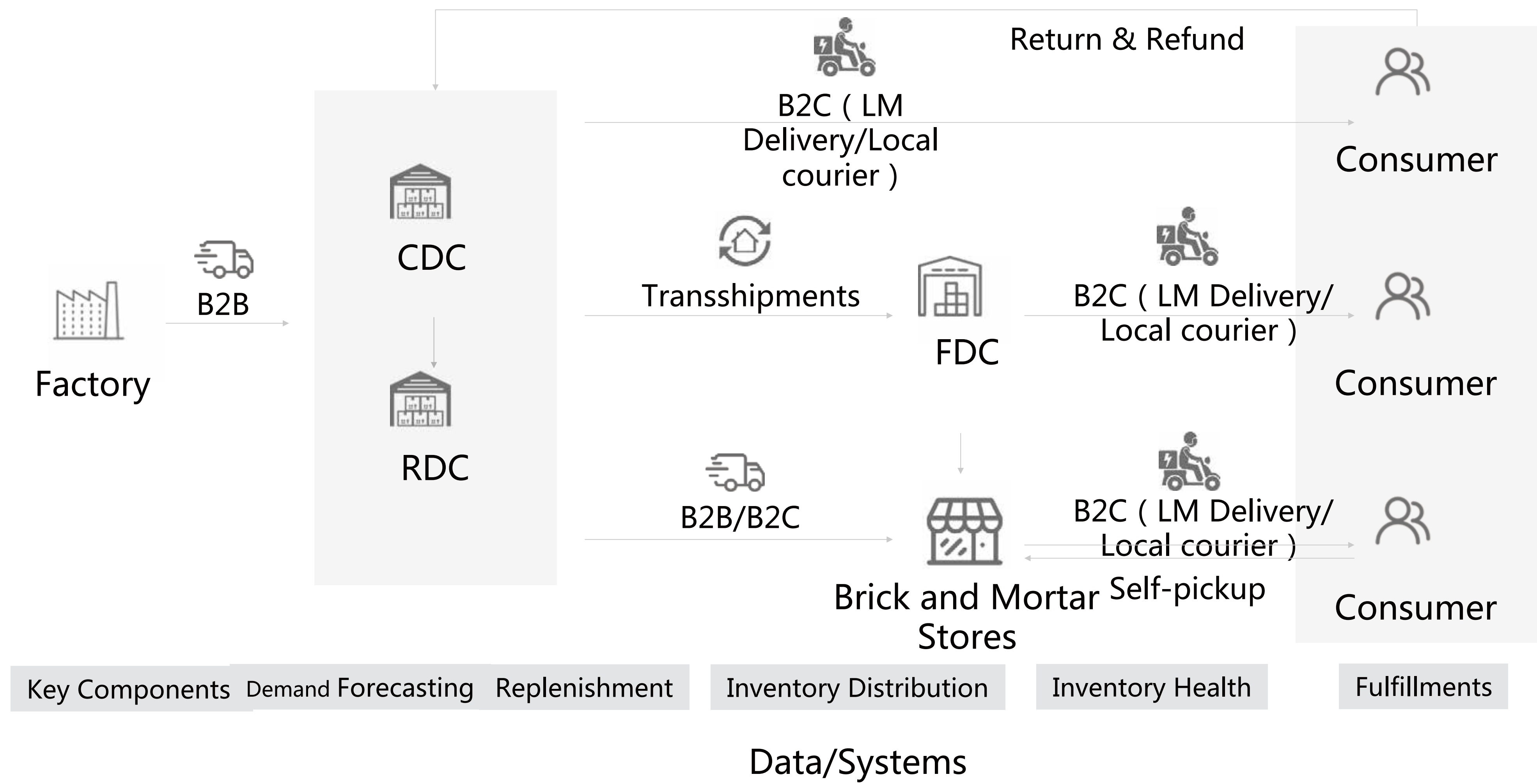
概述：

Information
Cash flow
Logistics

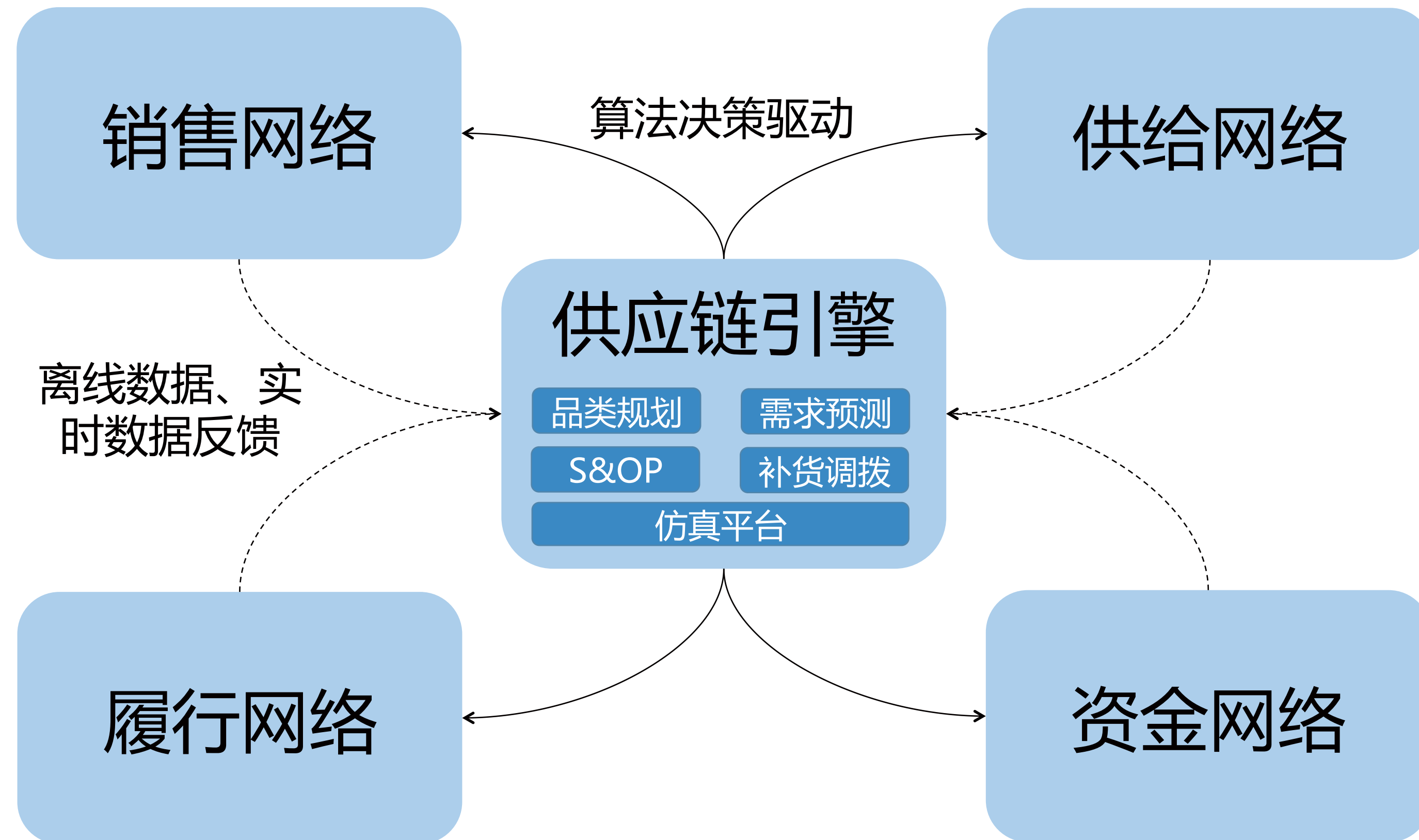


B2B2C Retail Supply Chain

供应链的关键组成成分：



供应链四大网络&供应链引擎：



1

概述

.....

2

销售网络

.....

3

供给网络

.....

4

履行网络

.....

5

总结&展望

销售网络-品类规划：

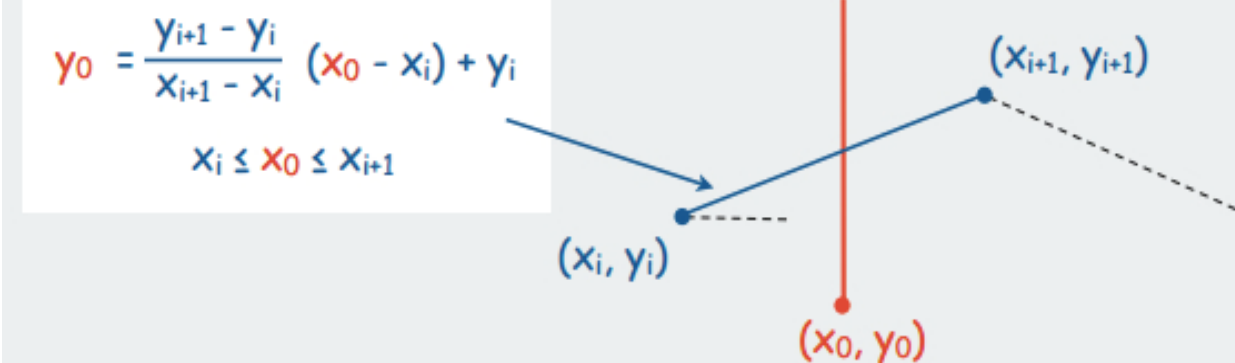
新品引入

- 采用深度网络对文本信息进行分析，提升总体精确度

$$\hat{d}_t(p) = \sum_{i=1}^N \frac{\mathcal{K}(w, w_i) \hat{d}_{ti}(p)}{\mathcal{K}(w, w_i)},$$

线下小店
选品

- 结合LBS数据、线上/下消费和行为数据，考虑订单满足率、渗透率及购买频次等约束，采用运筹优化算法进行智能选品



货品择仓

- 根据仓容约束、商品覆盖率约束和商品更替数量等约束，构建整数规划模型和启发式算法进行新仓选品，以提高子仓的覆盖率和减少拆单率

目标:

$$\text{minimize } w_1 \left(N - \sum_{i \in I^{new}} x_i \right) + w_2^+ x_2^+ + w_2^- x_2^- + w_3 \sum_{i \in I} x_i$$

约束条件:

$$\begin{aligned} y_o &\leq x_i \quad \forall i \in I(o), \forall o \in O \\ N(o) * y_o &= \sum_{i \in I(o)} x_i \quad \forall o \in O \\ \sum_o y_o + x_2^+ - x_2^- &= C \end{aligned}$$

商品分层

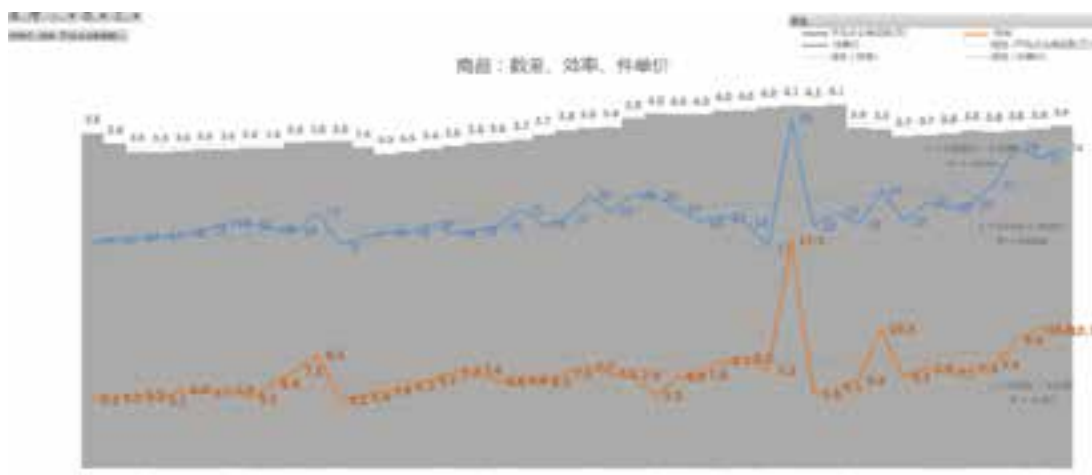
- 据GMV、高端人群占比、转化率、UV等指标圈定爆款、浅爆、长尾等商品分层，针对不同的分层采用不同的营销策略

$$c^{(i)} := \arg \min_j \| x^{(i)} - u_j \|^2$$

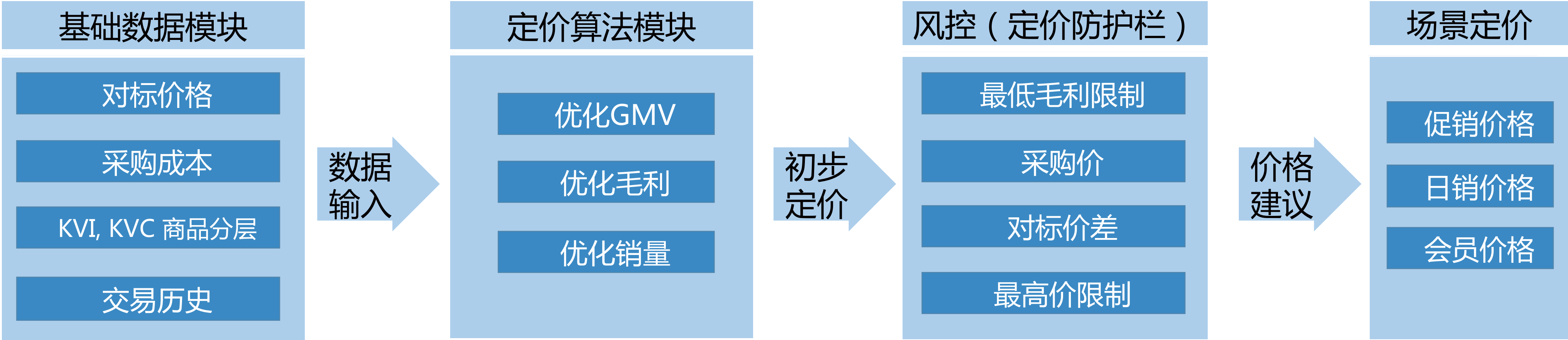
$$u_j := \frac{\sum_{i=1}^m 1\{c^{(i)} = j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)} = j\}}$$

汰换清仓

- 通过商品的季节性预测，协同营销节奏和权益，圈定清仓长尾滞销商品和出季商品



销售网络-动态定价策略：



货品分层	特点	目标	策略
KVI商品	高频率购买，高销量，高GMV	价格竞争力为主要目标	1. 确保价格有绝对竞争力
KVC商品	经常购买，高GMV		1. 满足一定利润要求 2. 确保价格有较高竞争力
普通商品	较低价格敏感度，较低交易量，较低GMV		1. 尽可能的提高利润空间 2. 价格有一定竞争力
	低价格敏感度，低交易量，低GMV	提升利润为主要目标	1. 尽可能提升整体利润 2. 价格可以较高

1

概述

.....

2

销售网络

.....

3

供给网络

.....

4

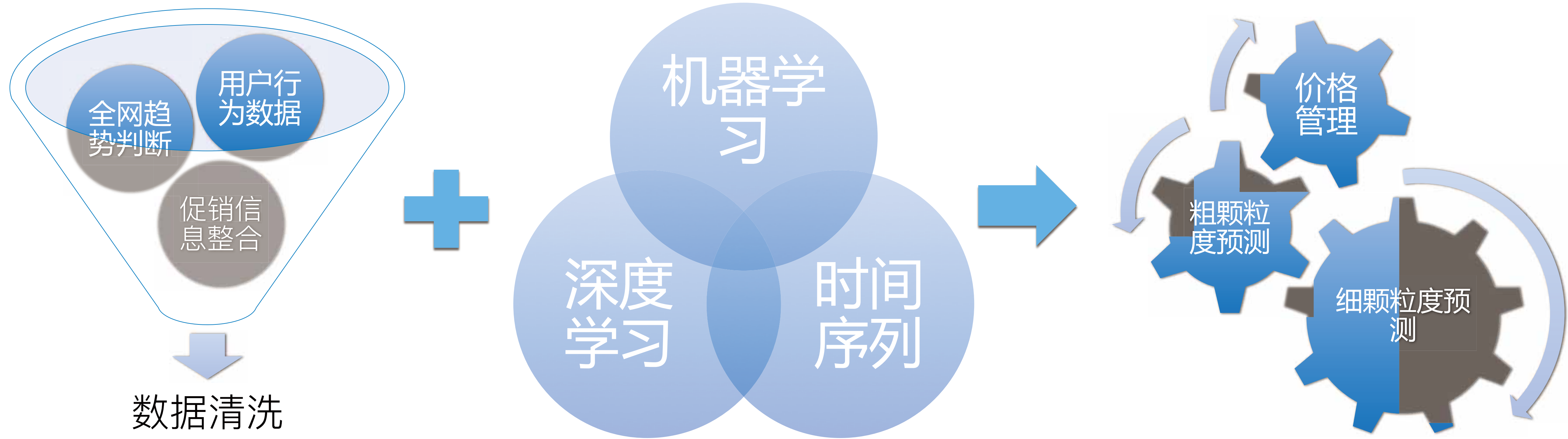
履行网络

.....

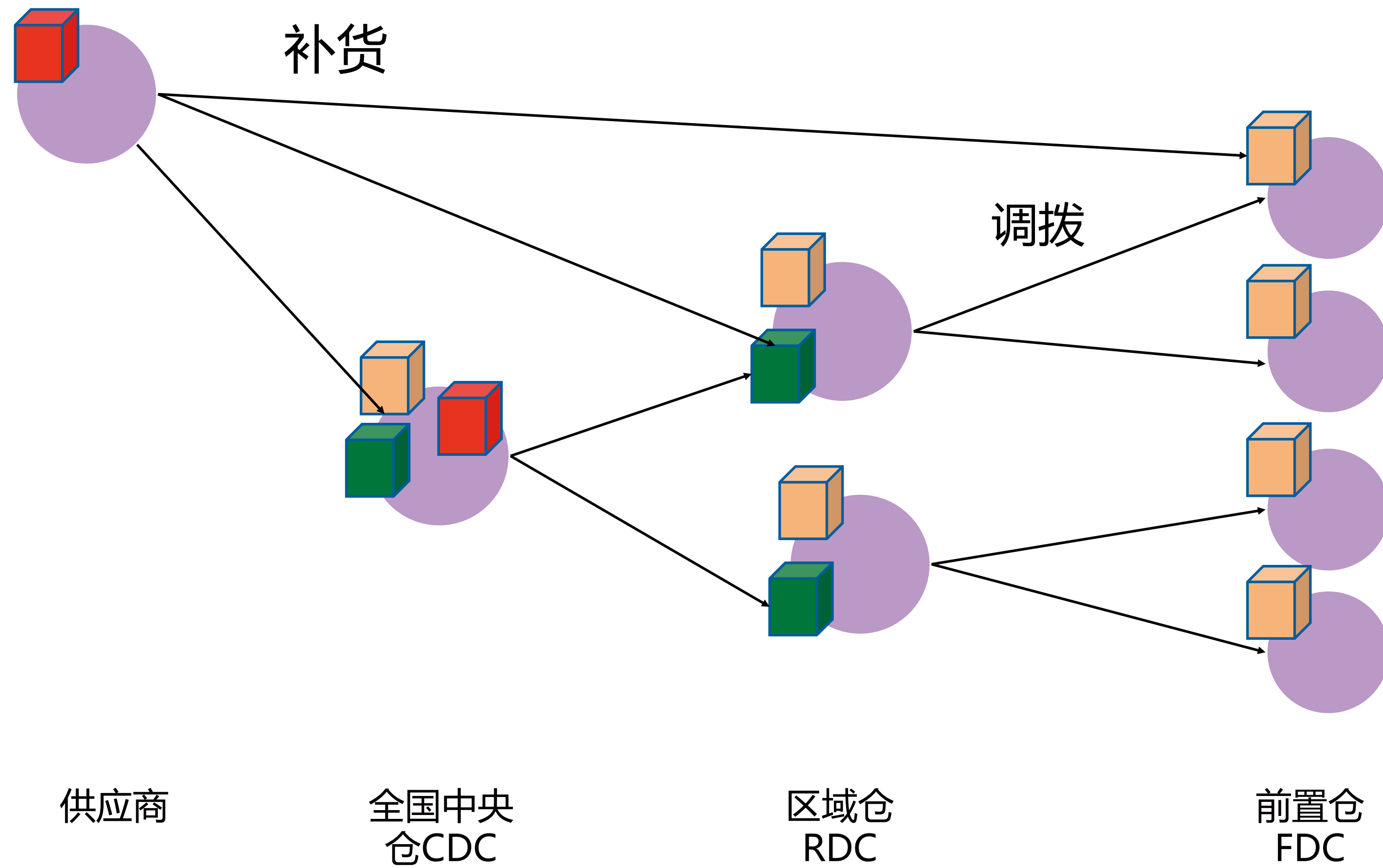
5

总结&展望

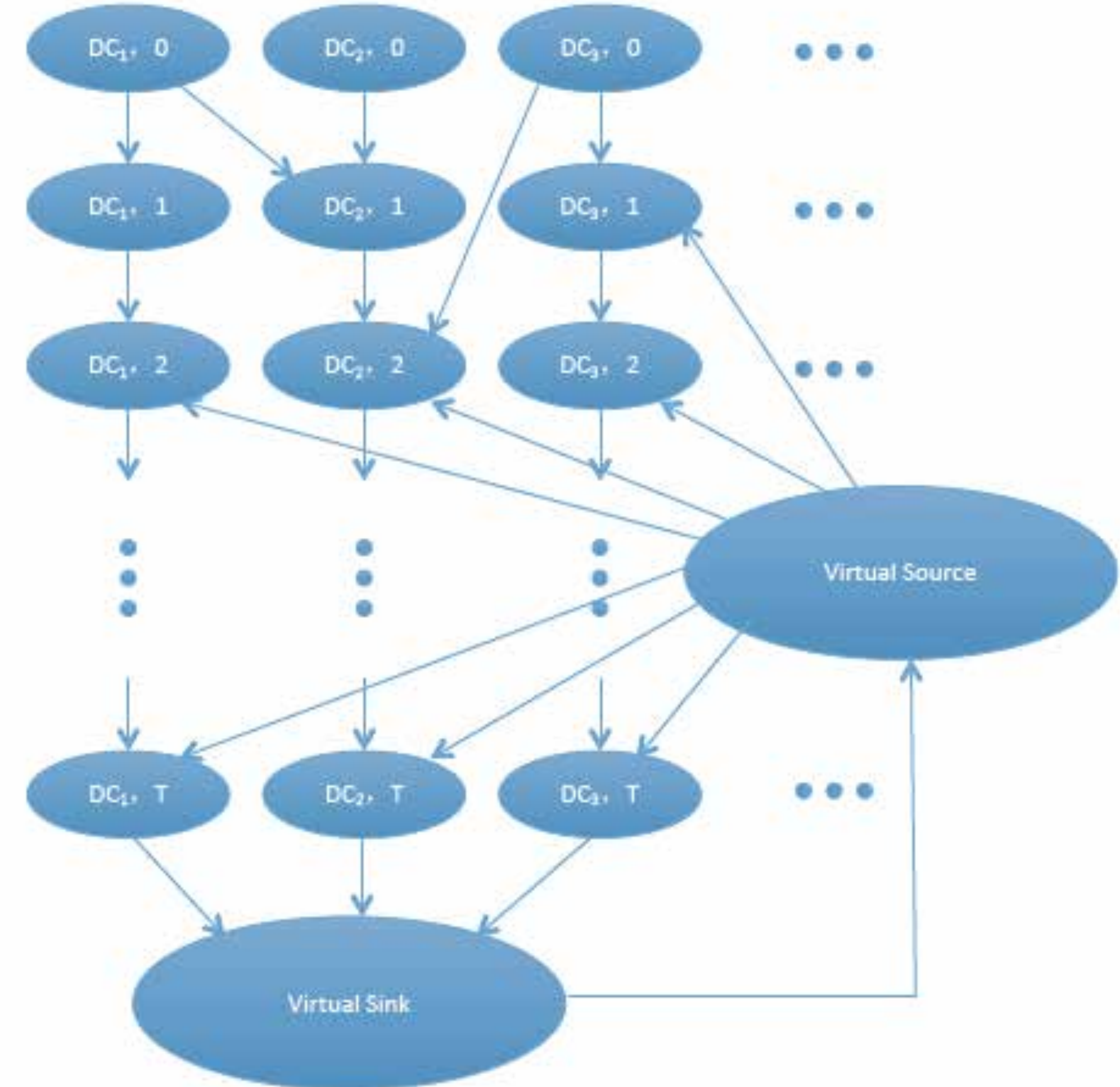
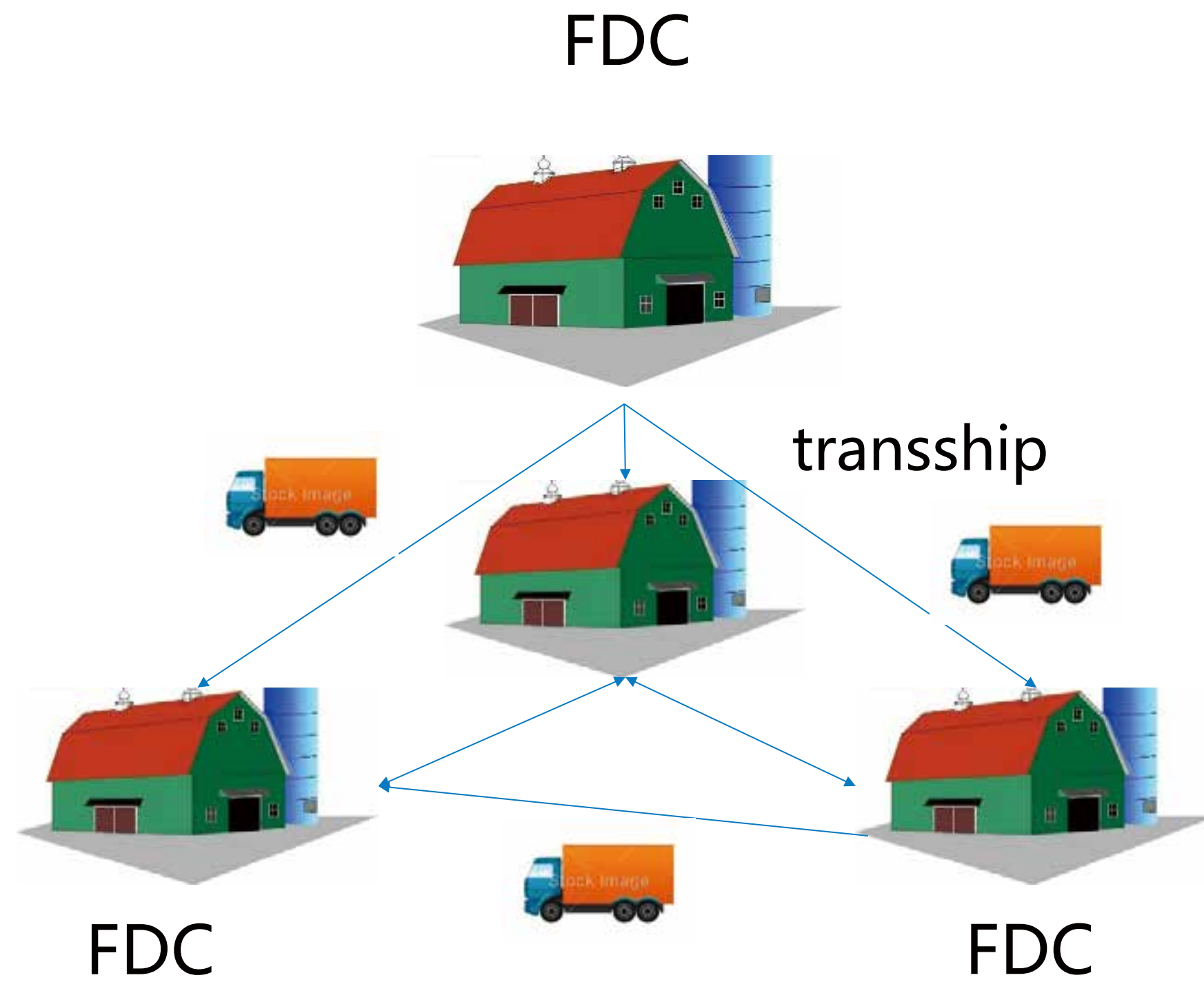
需求预测算法：



供给网络-多种补货方式：



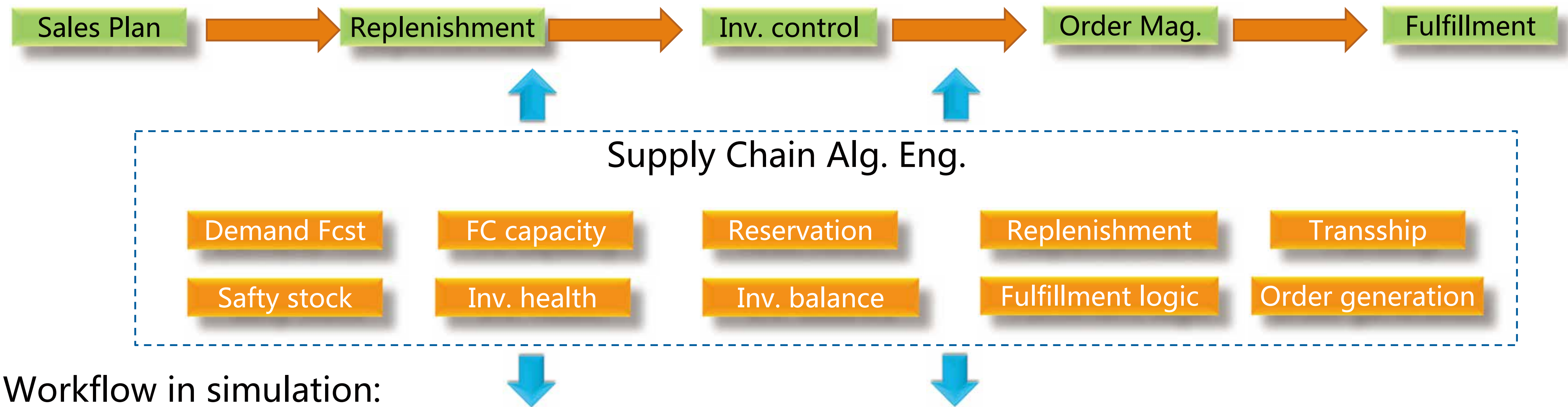
供给网络-调拨&库存平衡：



供应链仿真：

- 供应链系统是一个复杂系统，包含了多重随机性和业务逻辑，没有数学闭合解
- 单个供应链模块的改变，将会影响整个供应链的上下游其他模块，最终将会影响到整个供应链的性能
- 如何提升供应链的整体效率需要通盘考虑
- 新上线供应链模块对整体供应链有何影响

Workflow in reality :



Workflow in simulation:



1

概述

.....

2

销售网络

.....

3

供给网络

.....

4

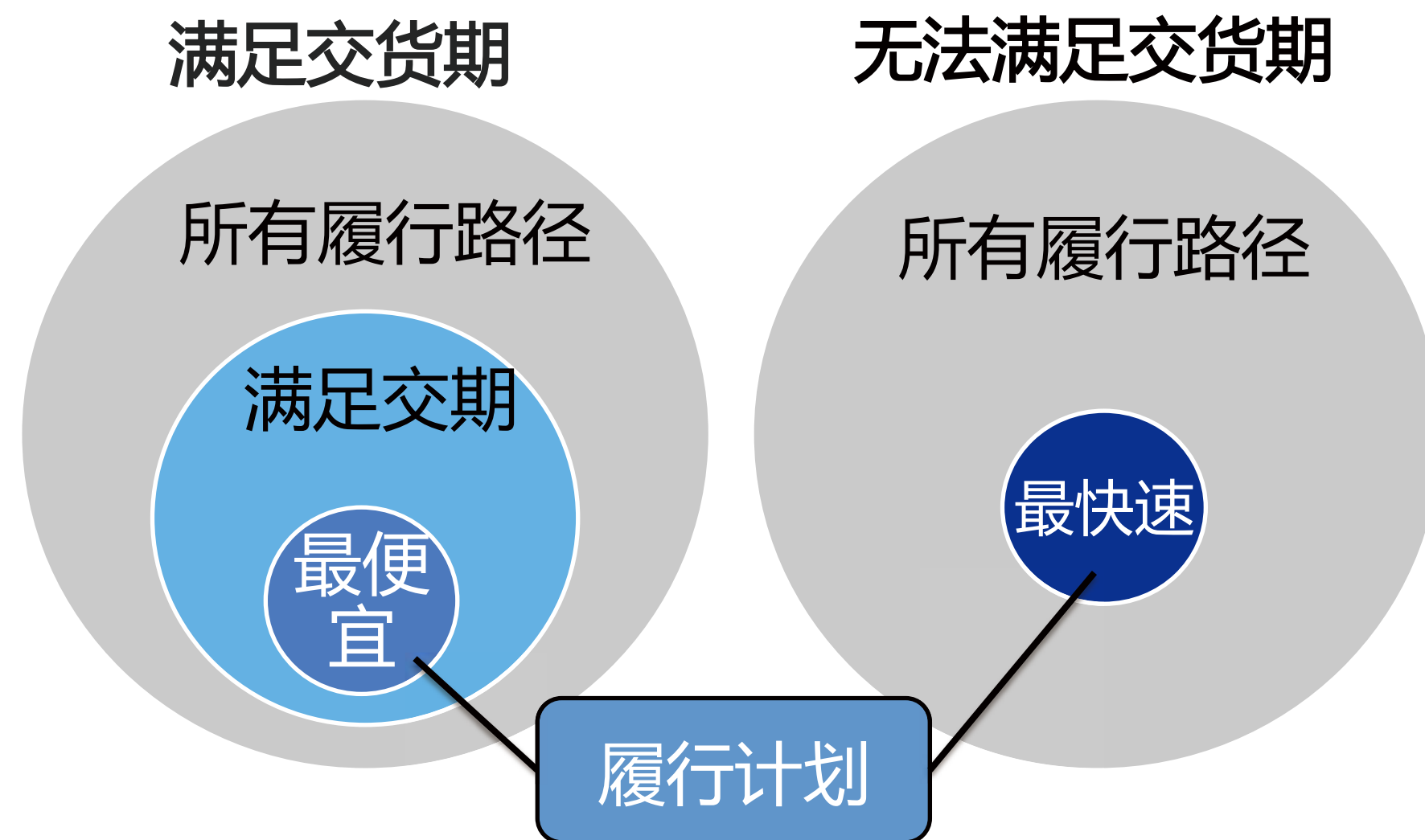
履行网络

.....

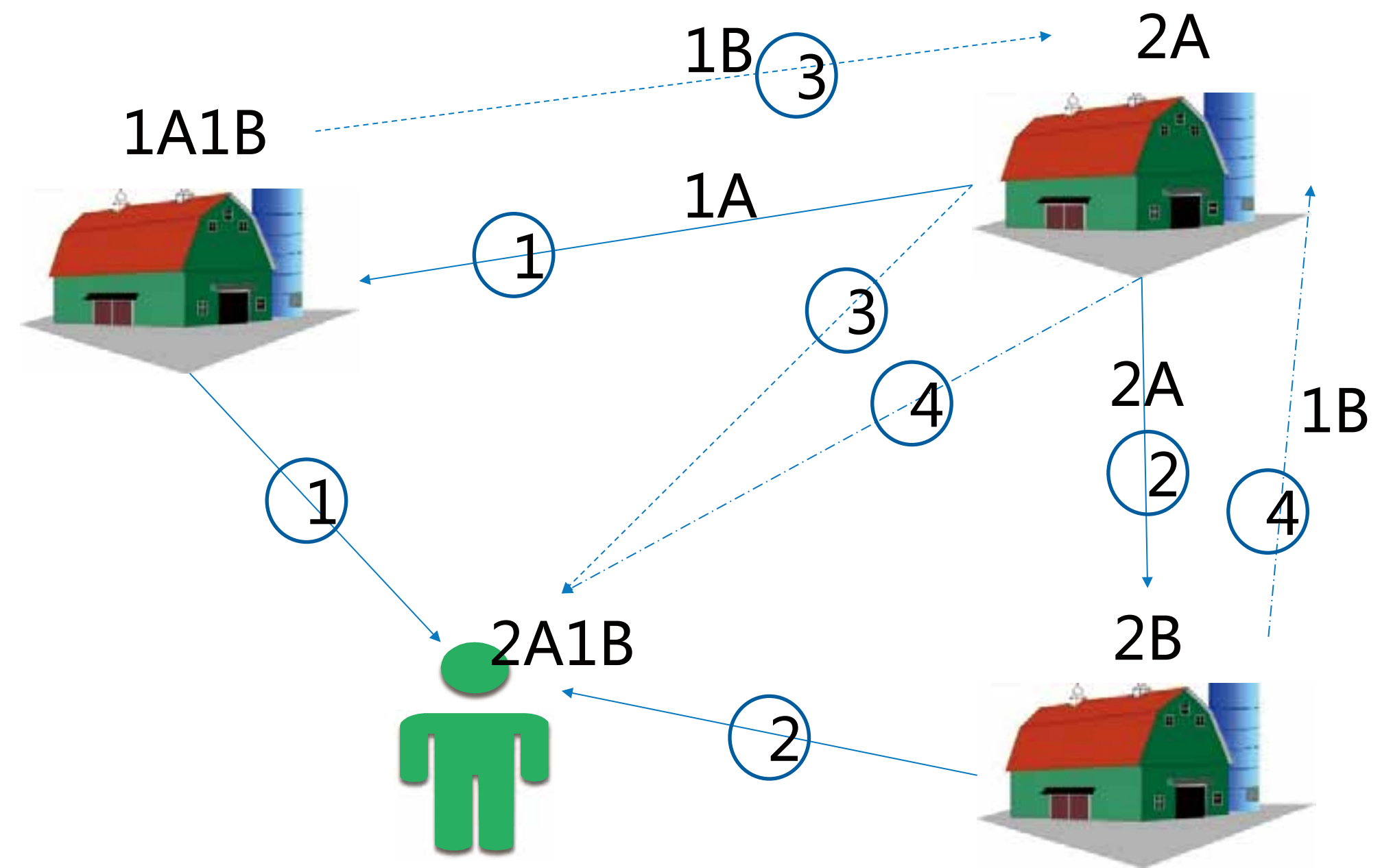
5

总结&展望

履行网络-选仓路由决策：



一个例子：



1

概述

.....

2

销售网络

.....

3

供给网络

.....

4

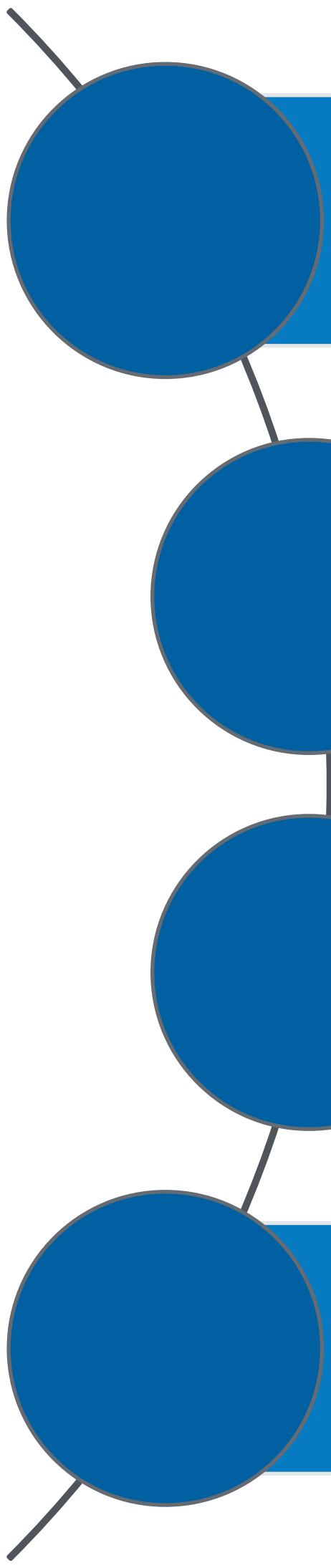
履行网络

.....

5

总结&展望

总结&展望：



供应链的复杂度增加，如何更好的去平衡需求与供应？

以大数据、IoT，社交网络为代表有革命性信息技术深刻的影响着供应链，如何使用这些数据与工具来改善供应链？

个人经验 VS. 机器学习

供应链对实时反馈和智能化的要求

走进蚂蚁金服技术

蚂蚁金服资深技术专家 南笑

Part 1

蚂蚁的诞生

蚂蚁的诞生

2014年10月蚂蚁金服诞生

蚂蚁：

从小微做起，也只对小微的世界感兴趣；像蚂蚁一样，虽然渺小，但是它们齐心协力，焕发惊人力量，在通往目的地的道路上永不放弃。

金服：

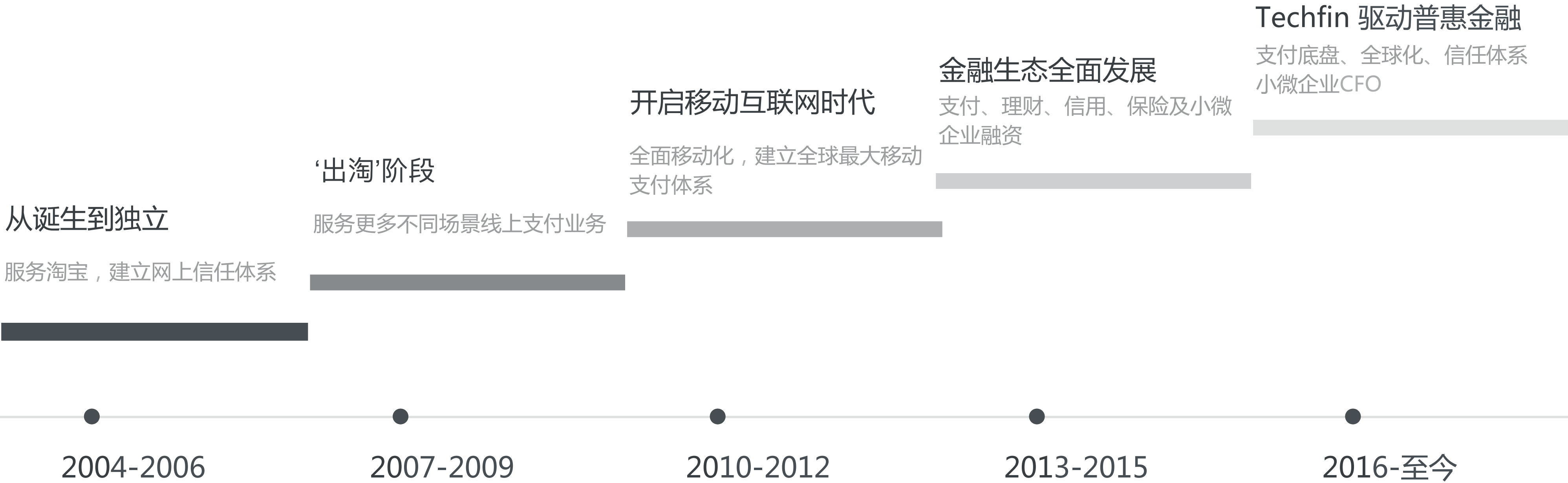
金服二字重在服务。

希望既能服务“草根”消费者和小微企业，也能服务金融机构，并和后者共同为未来社会的金融服务提供支撑。



蚂蚁的诞生

蚂蚁金服发展历程



Part 2

我们的梦想

无现金社会

绿色食品

无明矾油条

北京时间

23:36

浙江省
杭州市

信用即财富

黄慧 35岁 服装店主

“做生意难免有急用钱的时候，最怕欠钱又欠人情债。现在好了，凭信用就能借到款，半小时就到账。”

基于芝麻信用提供的征信服务，杭州银行推出“纯信用，随借随还”的信贷产品，帮助小微企业渡过难关。



蚂蚁金服
ANT FINANCIAL



杭州银行
BANK OF HANGZHOU

可持续环境



赋能全球合作伙伴



我们的梦想

好的金融:全球普惠金融



Part 3

我们所做的努力

我们所做的努力

The impossibilities today lead to the breakthroughs
tomorrow

技术存在边界，而边界指引我们探索和突破的方向

BASIC

Blockchain
AI
Security
IoT
Computing

Blockchain

区块链

AI

人工智能

Security

安全

IoT

物联网

Computing

计算

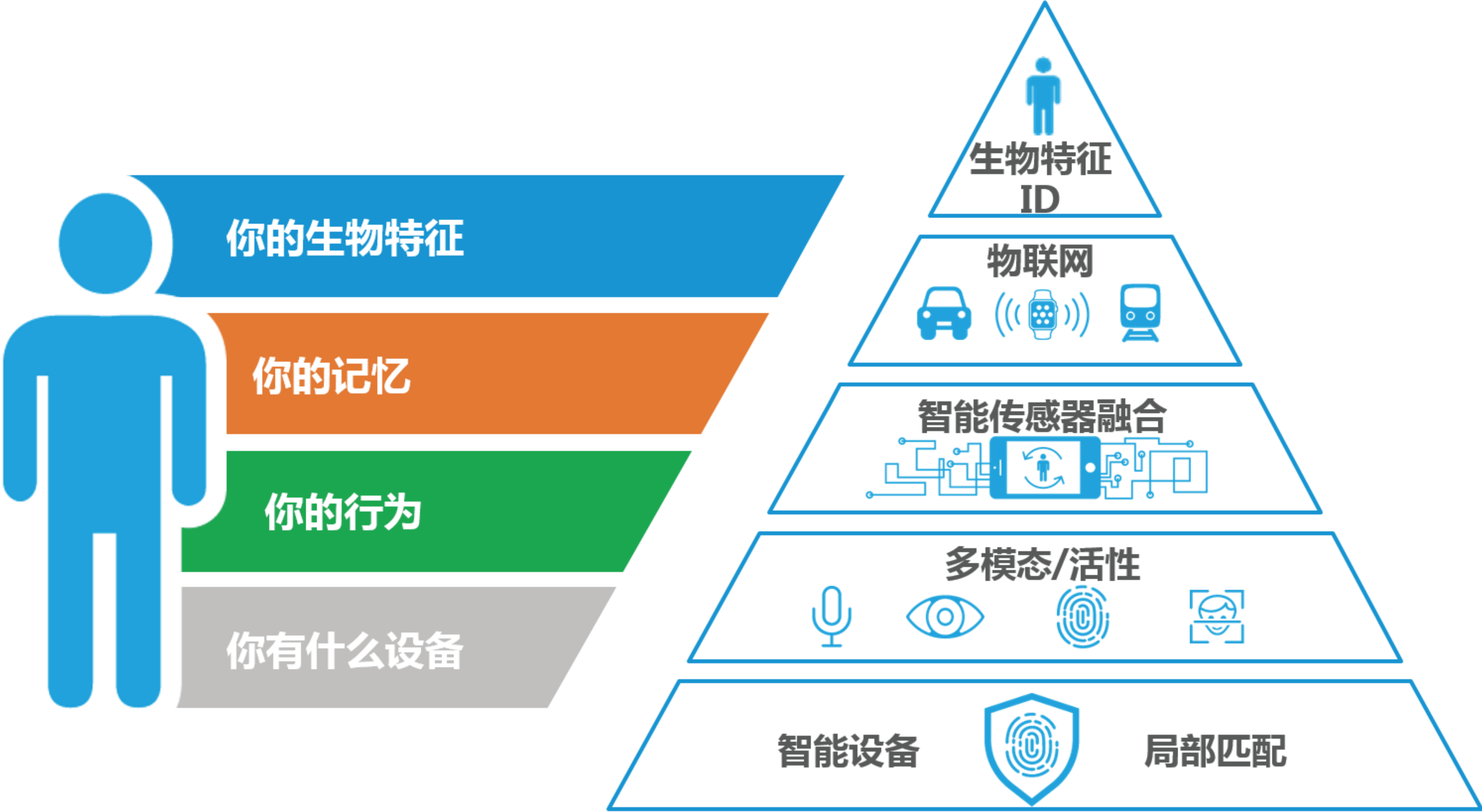
我们所做的努力

蚂蚁金服整体技术能力



我们所做的努力

生物识别技术



我们所做的努力

蚂蚁区块链-开放、透明的信任连接器



我们所做的努力

机器智能

NLP

机器学习

图像识别

语音识别

AI 平台

AI 应用



信用



推荐



风控



搜索



智能助理



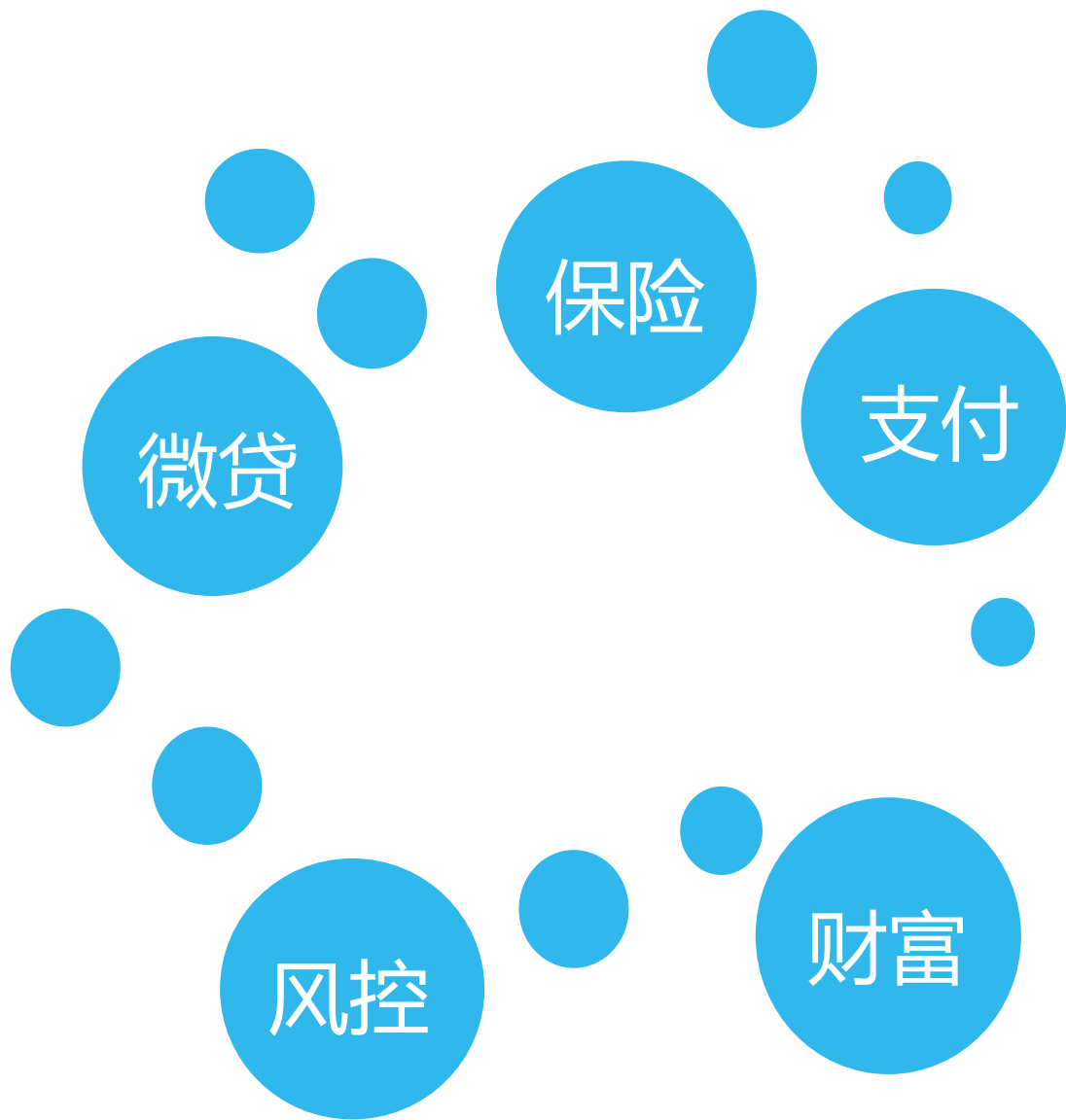
营销

Part 4

我们的挑战

我们的挑战

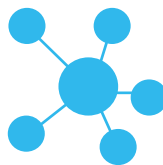
金融服务的基本挑战



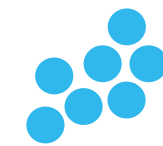
金融服务的
基本挑战



时间敏感



海量数据



业务多样性



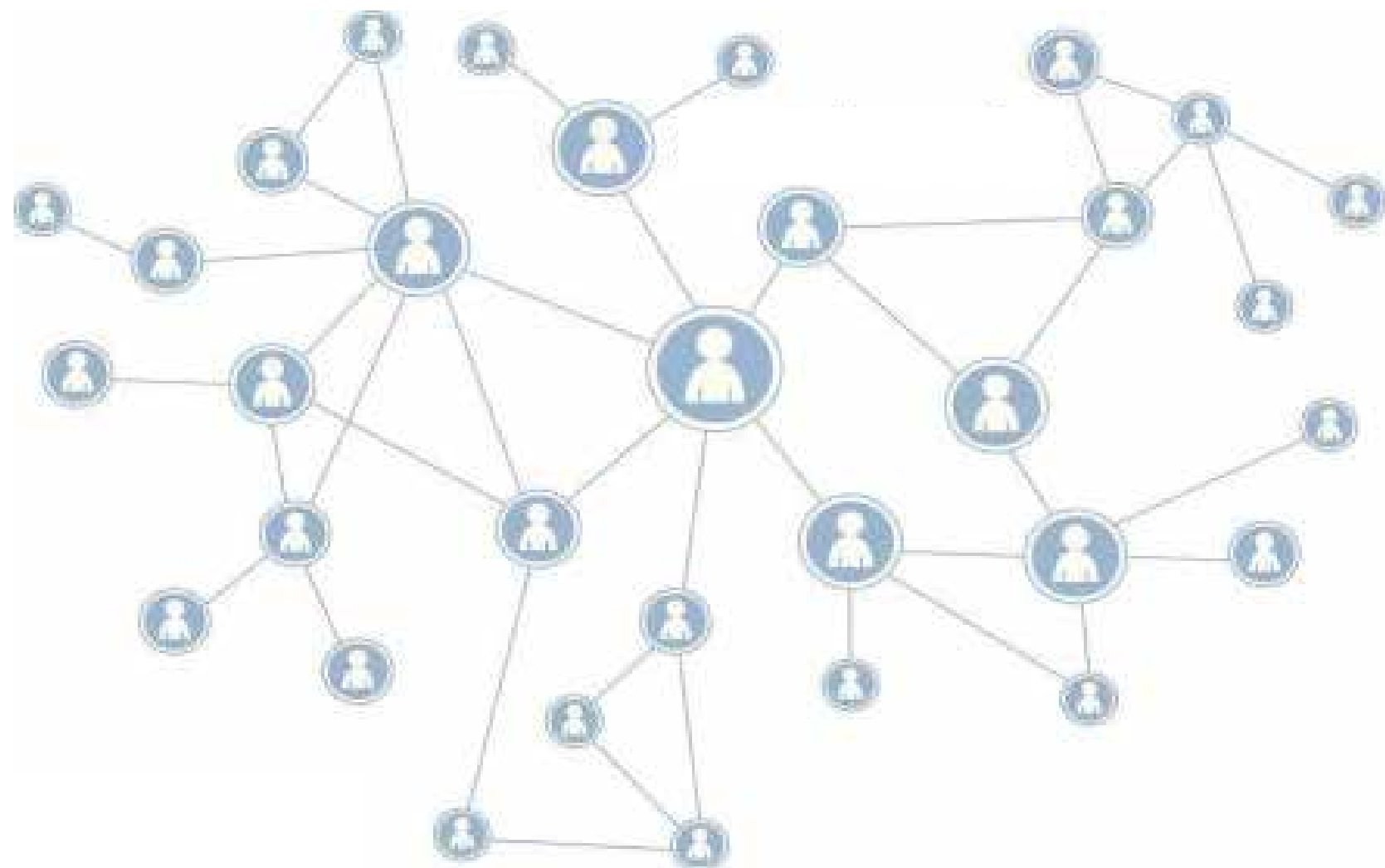
系统性风险



强安全

我们的挑战

智能安全



系统安全检测、反盗用、反洗钱、反欺诈、反作弊、反套现

系统安全检测

弱监督学习

发现潜在威胁

提升攻击检测

检测精度可达到94%

反洗钱模型

无监督或者弱监督图算法

资金交易网络特征挖掘

无监督隐案挖掘

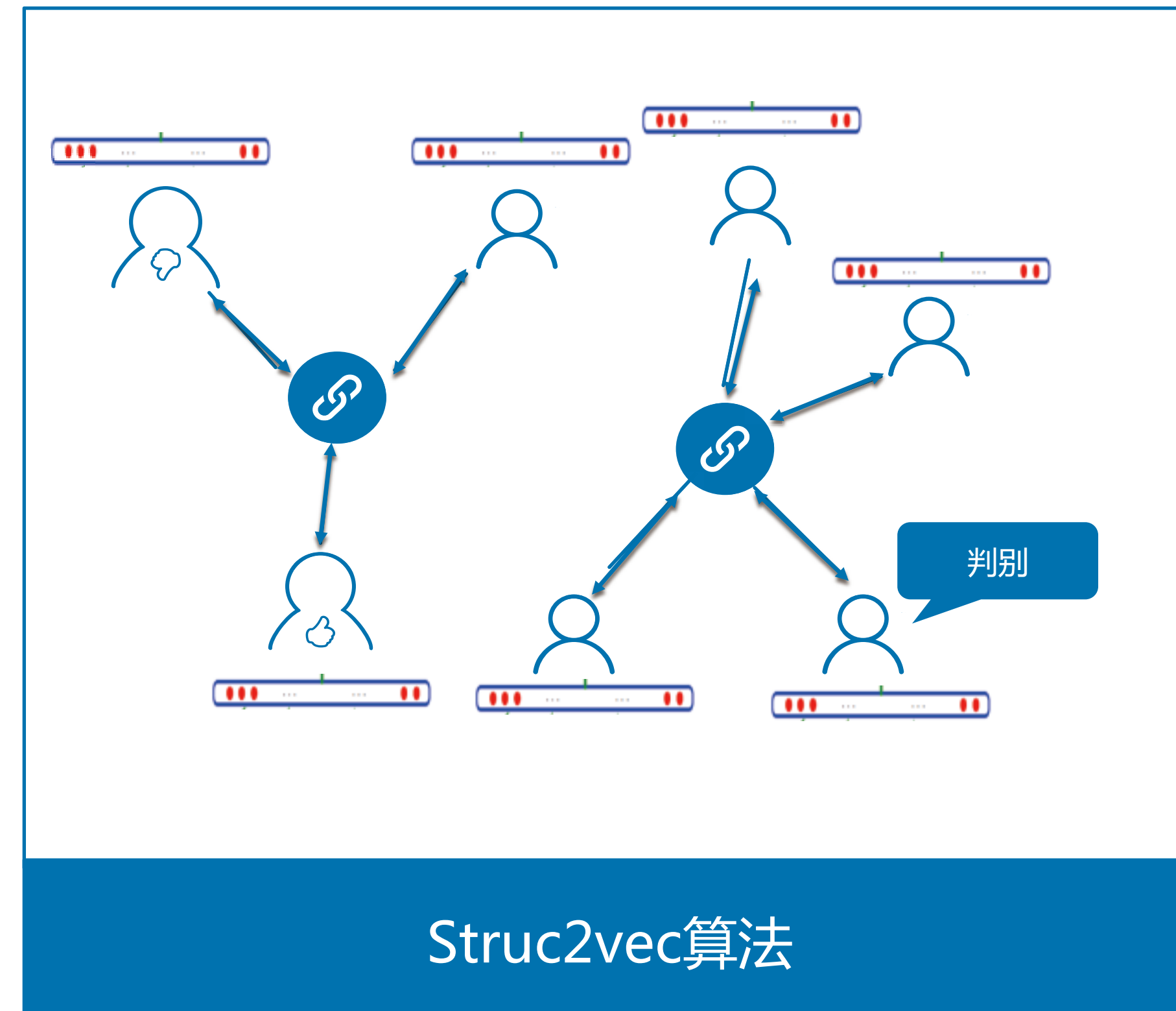
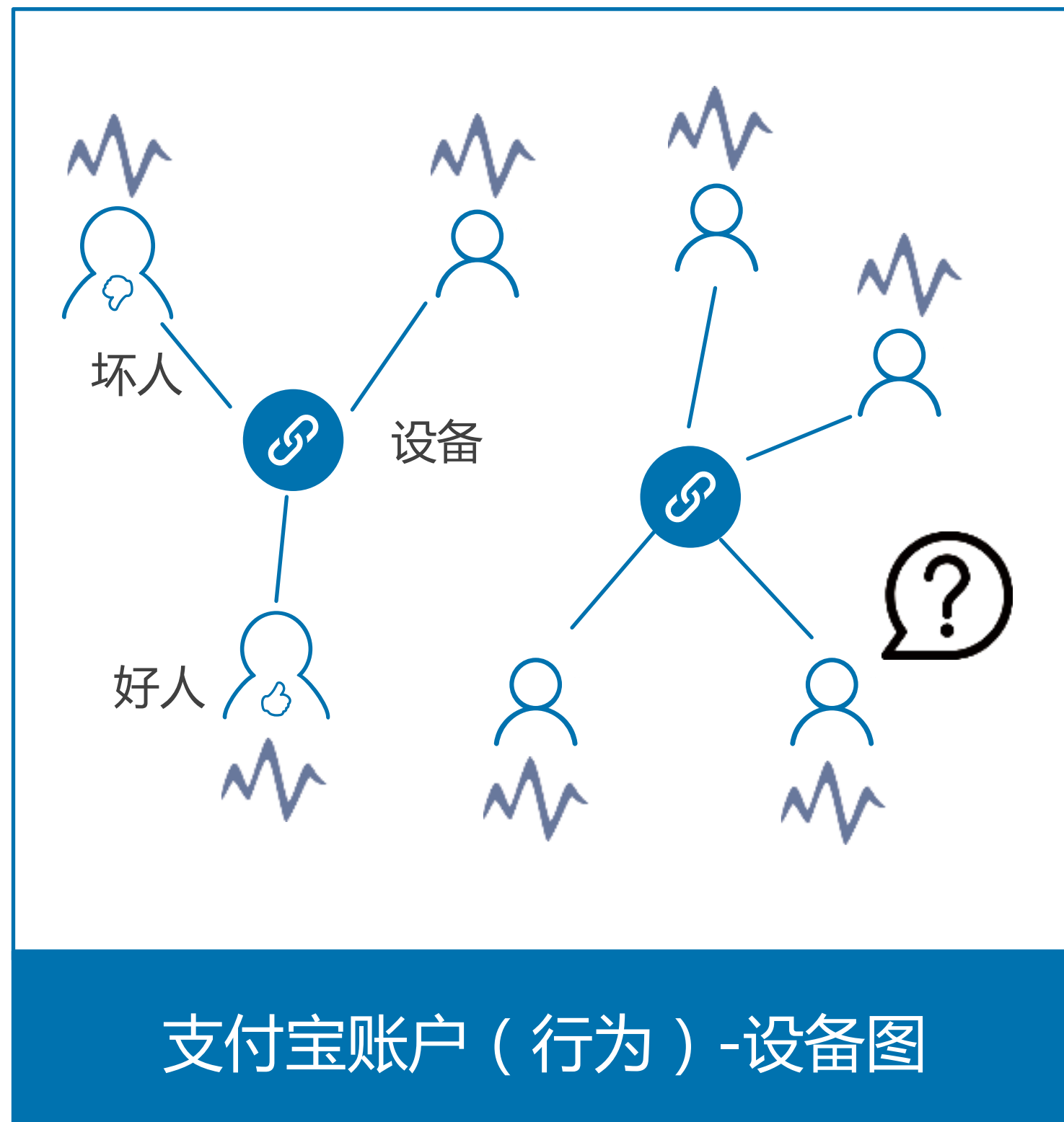
地下钱庄TOP-30可疑样本发现16个正确

我们的挑战

智能安全

价值：垃圾账户注册防控能够降低后端风险基数，稳定大盘指标，极大的提高整体的账户质量。

挑战：新注册用户缺少画像信息，如何准确判别一个新注册账户是不是垃圾账户？

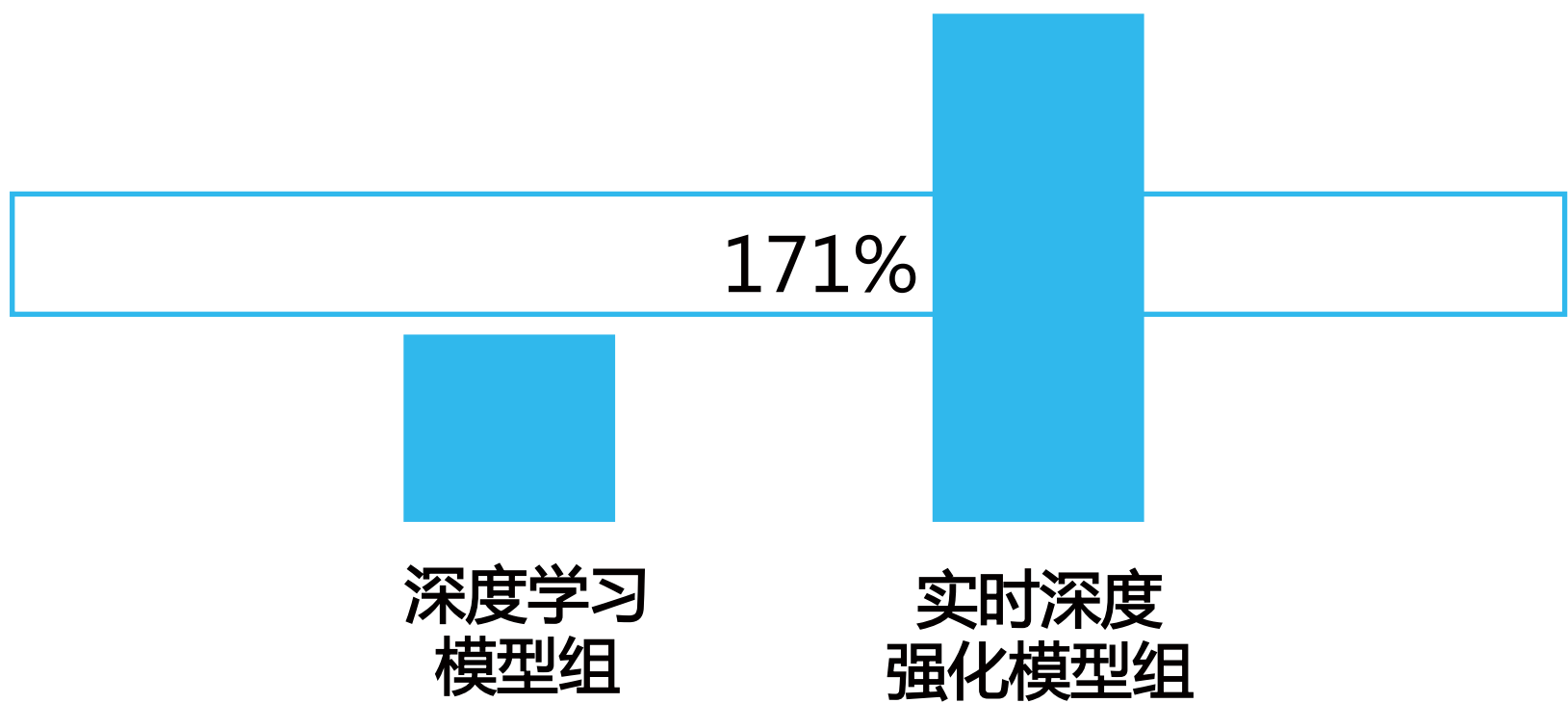


我们的挑战

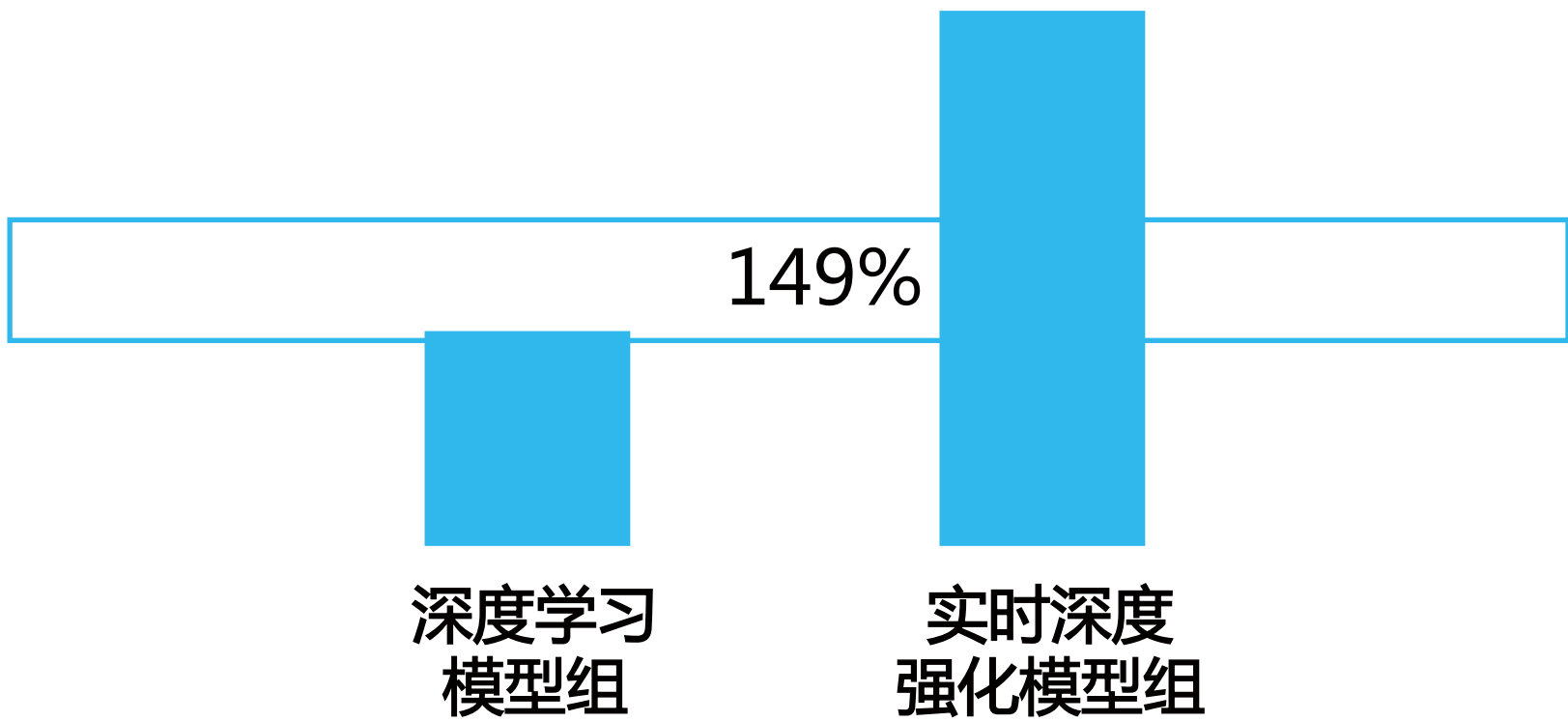
智能营销



推荐卡片点击率对比



最终签约率对比



我们的挑战

保险应用

智能定价

精准定损

降低成本

风险防范

极速核赔



退运费险

千人千面精准定价



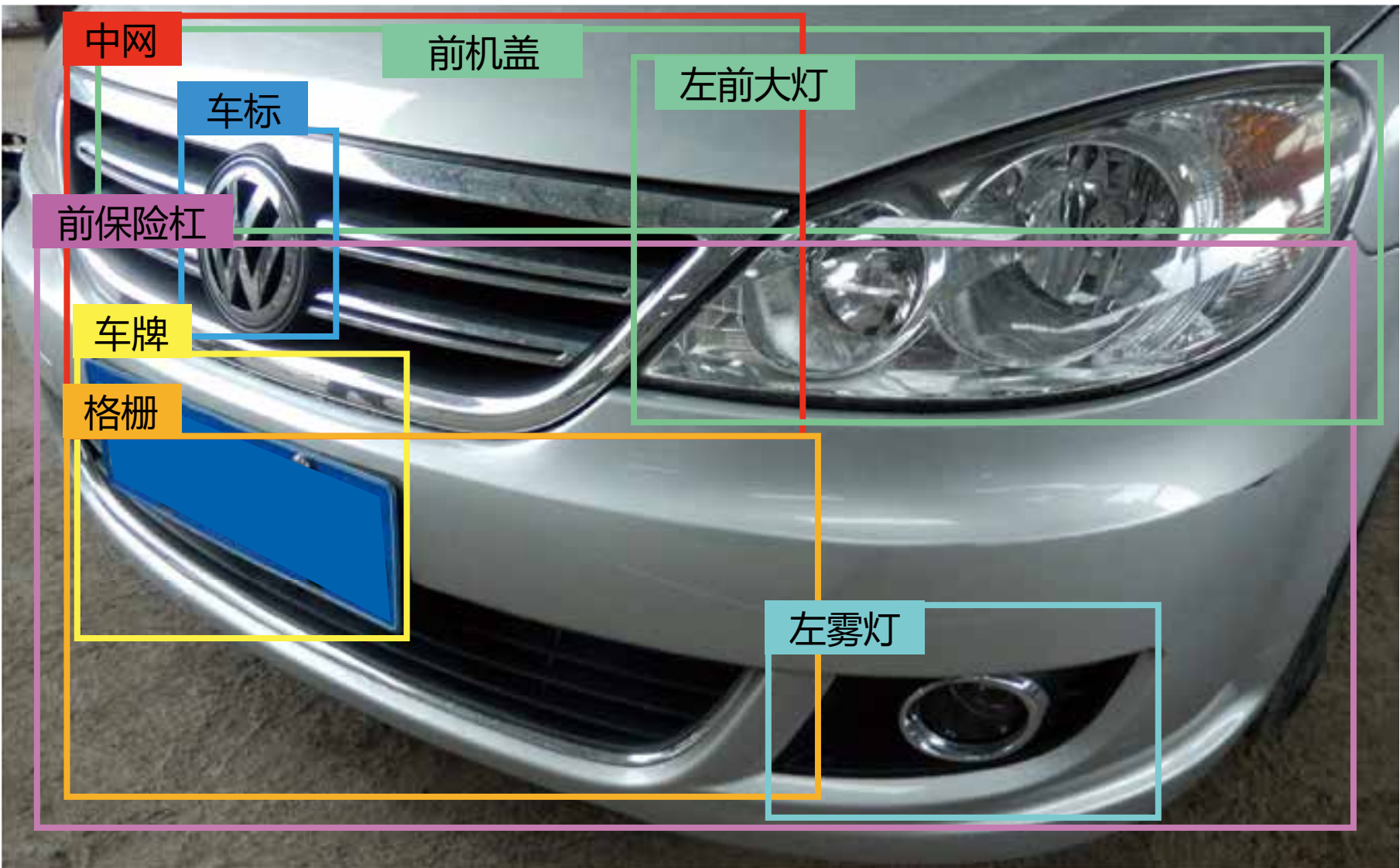
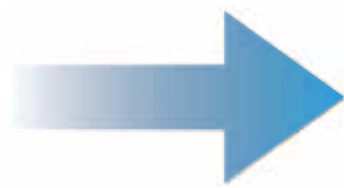
定损宝

相当于10年以上定损专家



在线理赔

在线报案>风控拦截>审核决策
理赔效率提升50%



我们的挑战

保险应用 – 定损宝

定损宝案例



鉴定结果：左侧后翼子板轻微损伤

赔付金额：500元

我们的挑战

Antone

懂你



帮你



多轮对话

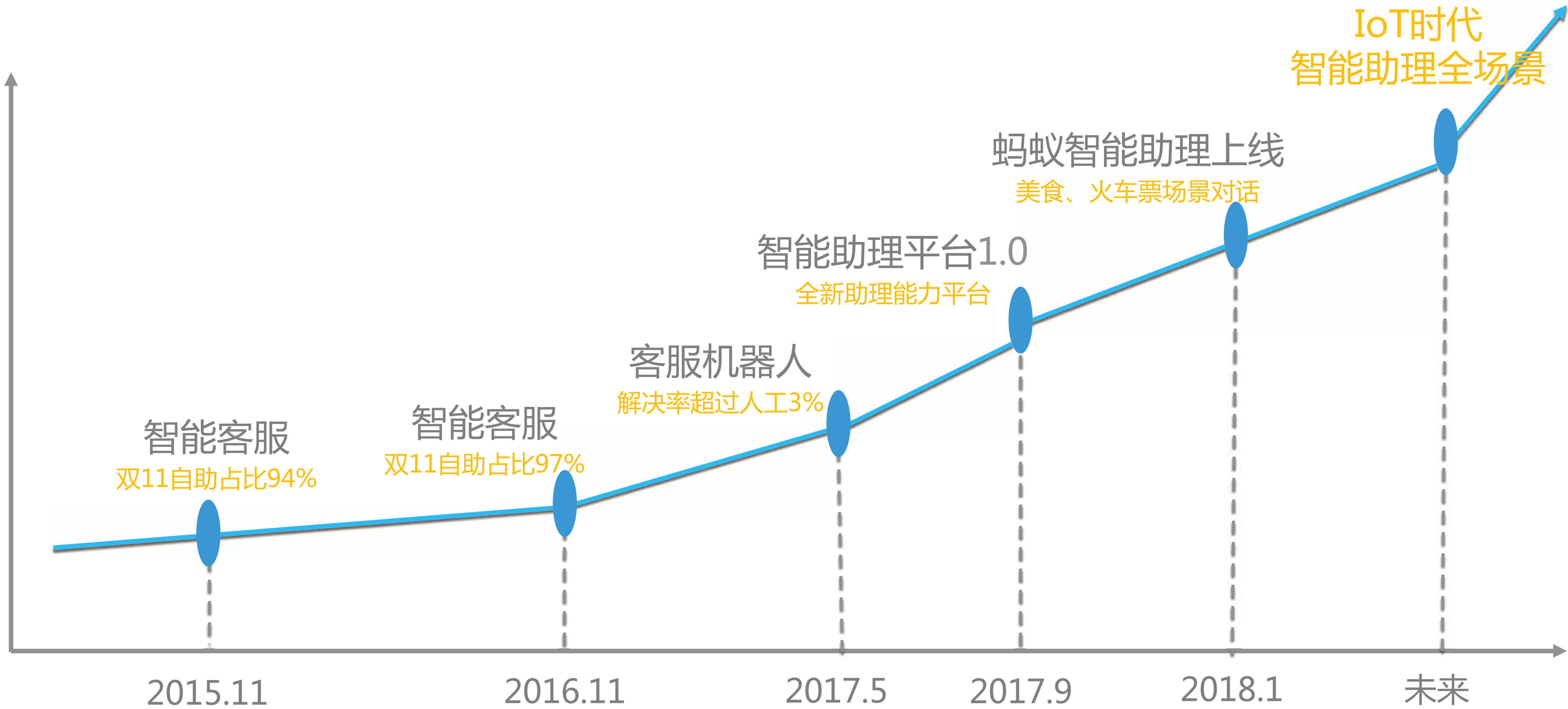
用户意图多轮交互
给用户一站式解决方案

业务直达

覆盖支付宝500+

我们的挑战

Antone诞生历史



我们的挑战

金融智能平台



安全

软硬件加密
数据安全



实时

低延迟在线预测
硬件加速



大规模

高吞吐离线训练
分布式实时计算

强化学习

深度学习

无监督学习

图推理

... ..

金融级超大规模异构计算

GPU

FPGA

CPU

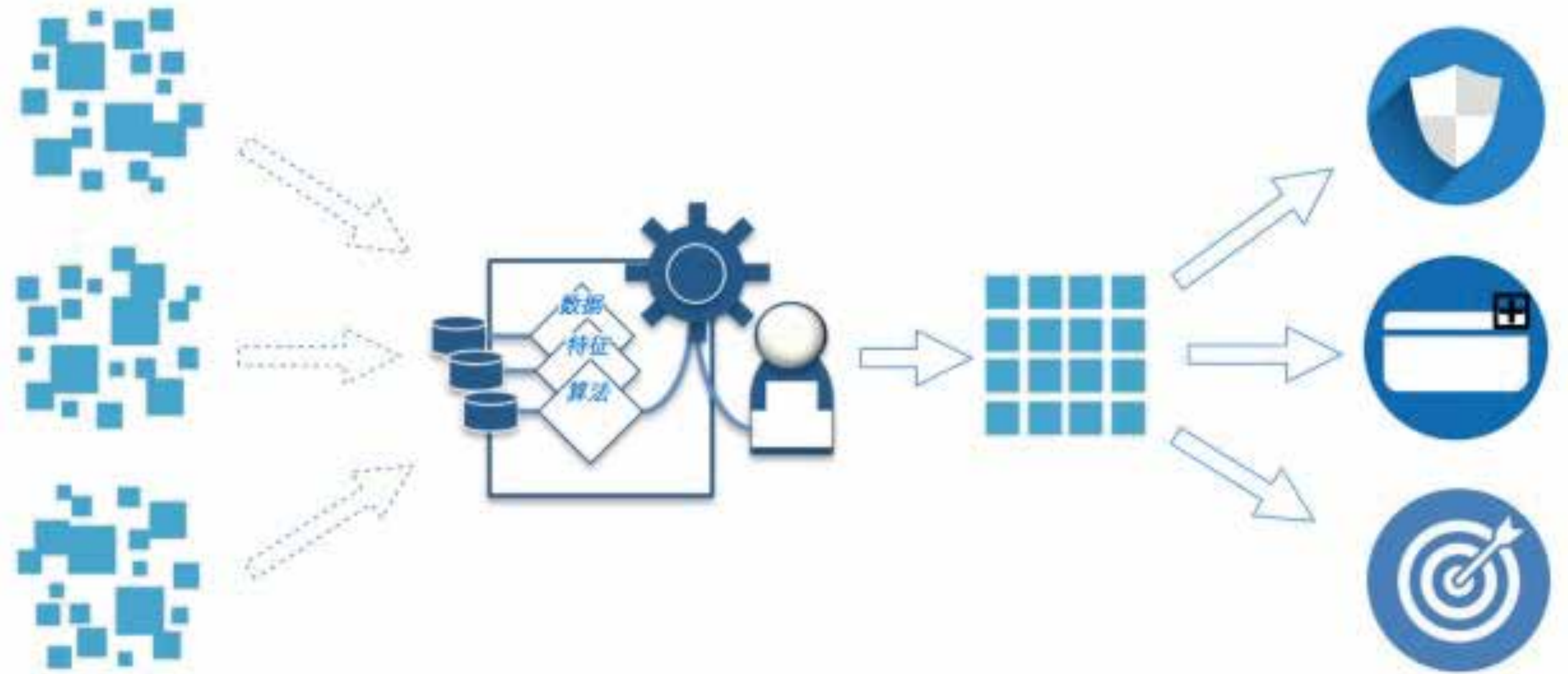
ASIC

... ..

我们的挑战

金融智能平台

- ✓ 可视化建模
- ✓ 自动化部署与模型服务
- ✓ 资产化沉淀
- ✓ A/B Testing
- ✓ ...



我们需要你！

菜鸟网络智慧物流： 算法应用的新蓝海

菜鸟网络人工智能部资深算法专家
元享

目录

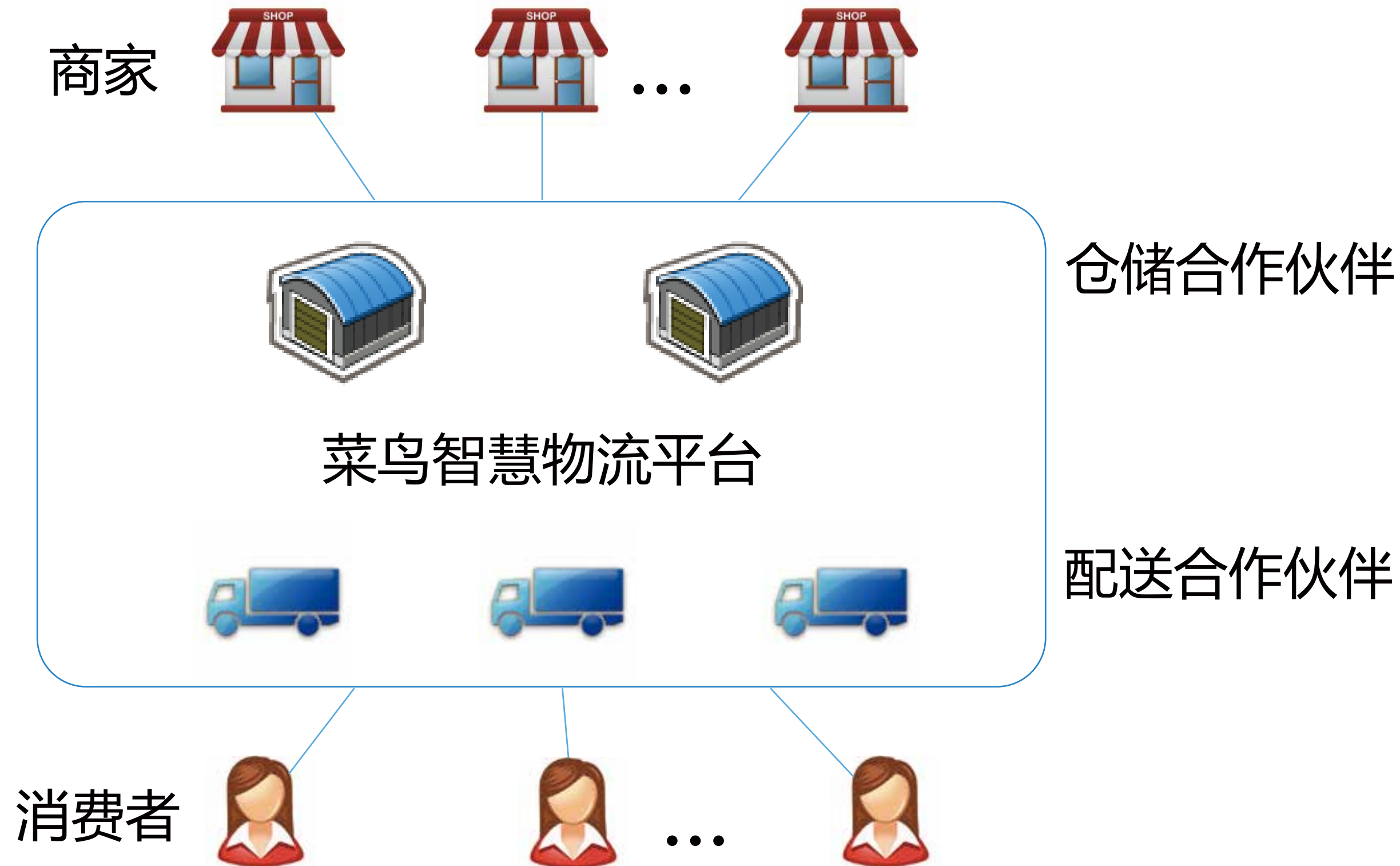


Part 1

菜鸟智慧物流介绍

菜鸟智慧物流平台

连接物流合作伙伴 + 数据&技术赋能



打造物流行业的基础技术设施

电子面单
地址库
物流云

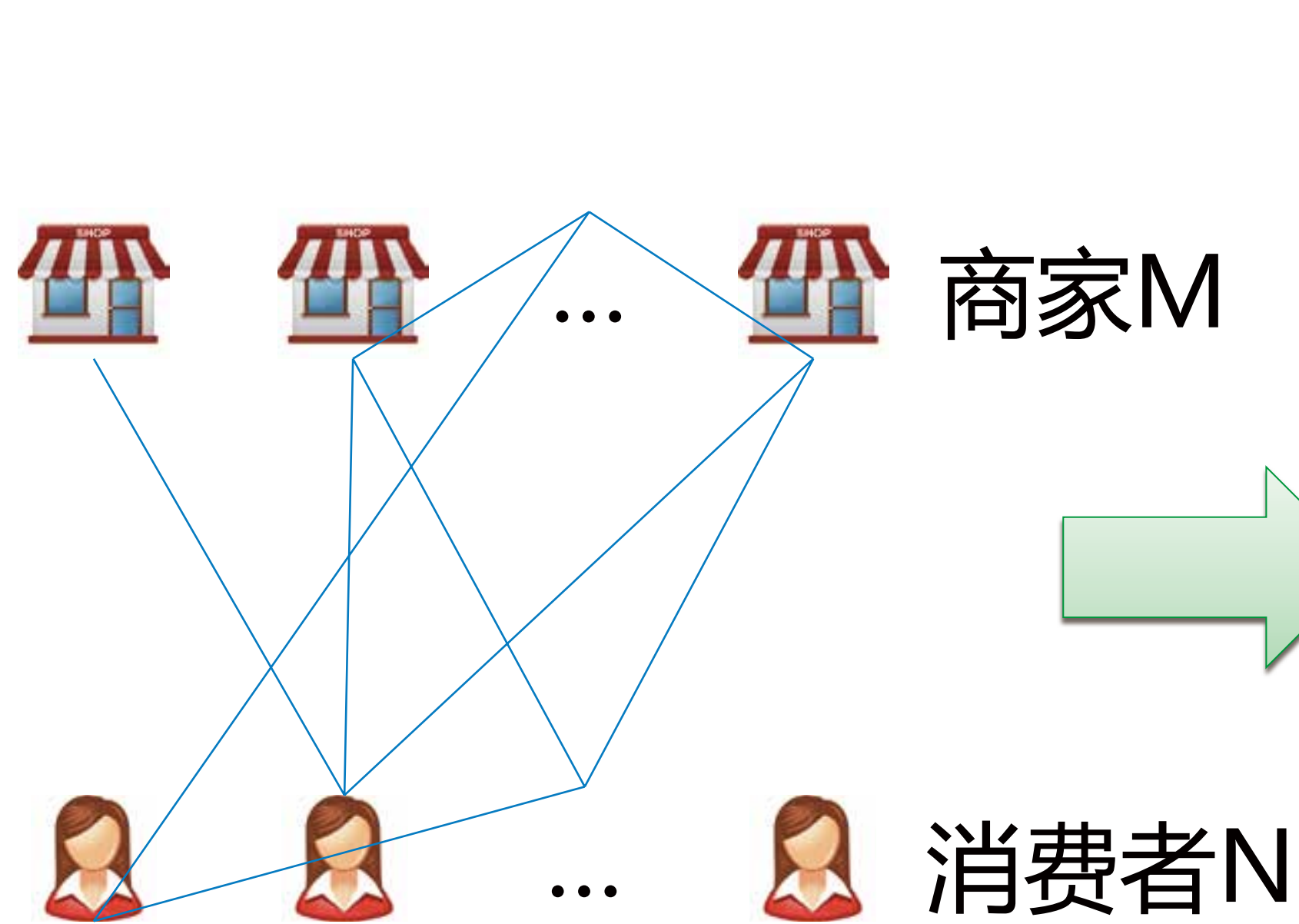
联合多个物流合作伙伴(CP)，协同作业

仓+干线+分拨+配送

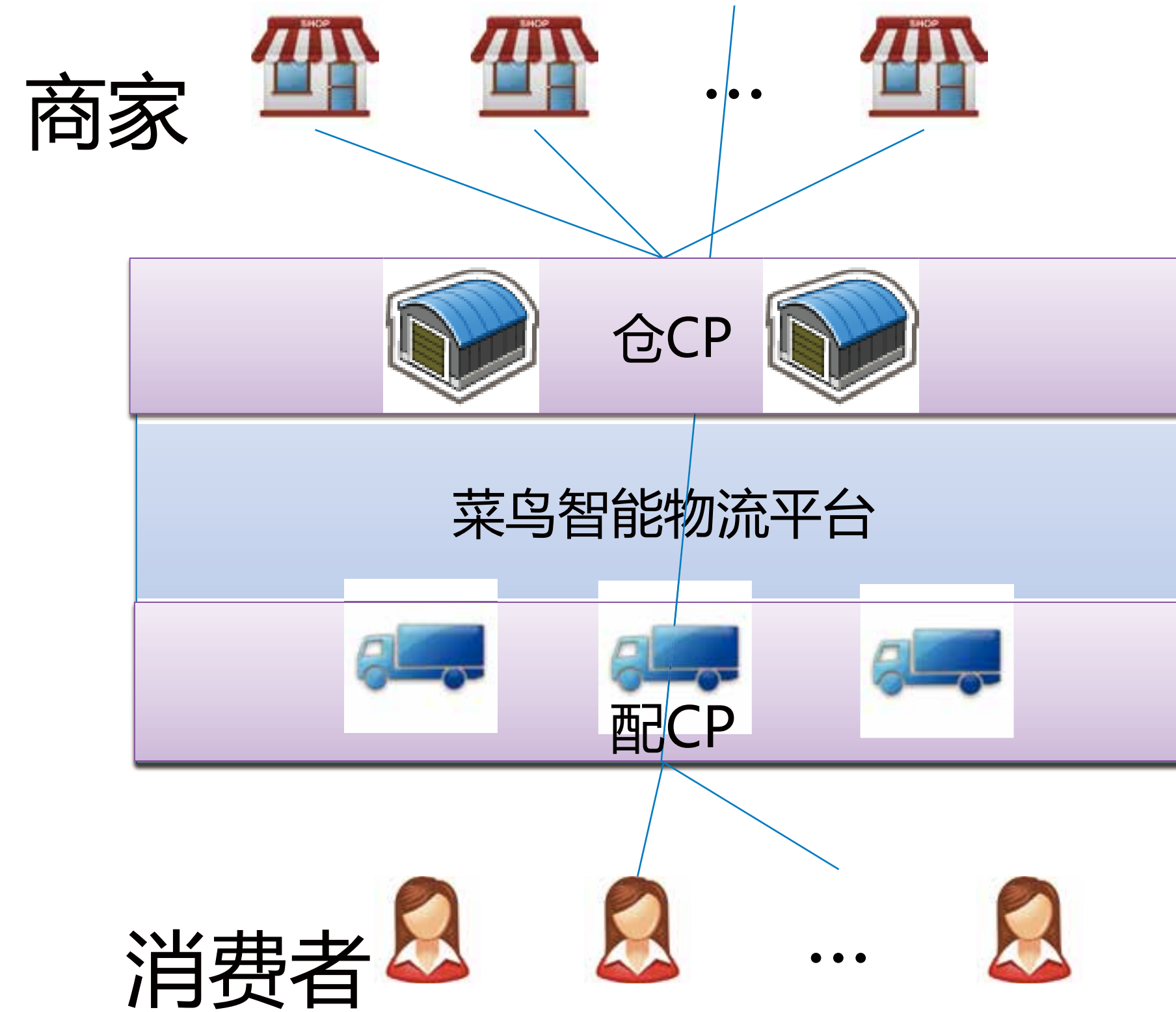
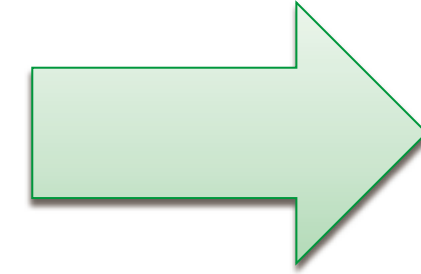
将合作伙伴的物流服务能力组合成网络和打包服务，提供给商家

仓配服务
快递服务
末端服务
跨境物流服务
农村物流服务

社会化协同平台



问题复杂度： $O(MN)$



问题复杂度： $O(M+N)$

复杂的事情让菜鸟来做

菜鸟网路现状

全国24小时
必达

全球72小时
必达

“未来商业” 的
基础设施



快递网络



仓配网络



跨境网络



末端网络



农村物流

菜鸟现状（截止2017年3月）

- 覆盖中国**70%**快递包裹数
- 货通**2700+**个区县
- 连接~**200万** 快递员
- **20+万**合作伙伴运输车辆

Part 2

菜鸟的技术特点

技术挑战 @ 智慧物流

算法模型层面

1. 实物流的机器学习问题：
uncertainty，可解释性
2. 组合优化问题: small data, large
search space
3. 链路长，Decision Processing
Optimization 问题，
4. 算法结果与实操的GAP
5. 机器学习和运筹优化算法的融合
6. 黑盒优化问题
7. 多目标，对复杂业务逻辑和模糊目标的理解
和建模
8. 实物流数字化建设（机器视觉&Iot）
9. 多智能体学习问题

工程技术层面

1. 计算复杂度高，实时
性要求强
2. 超大规模启发式优化
引擎的分布并行化
3. 解空间的合理化分
区，索引，更新
4. Offline graph
building，nearline
graph updating，
online decision
making的架构
5. 自动化牵涉到硬件软
件的多端协同

效果评估

AB test

仿真

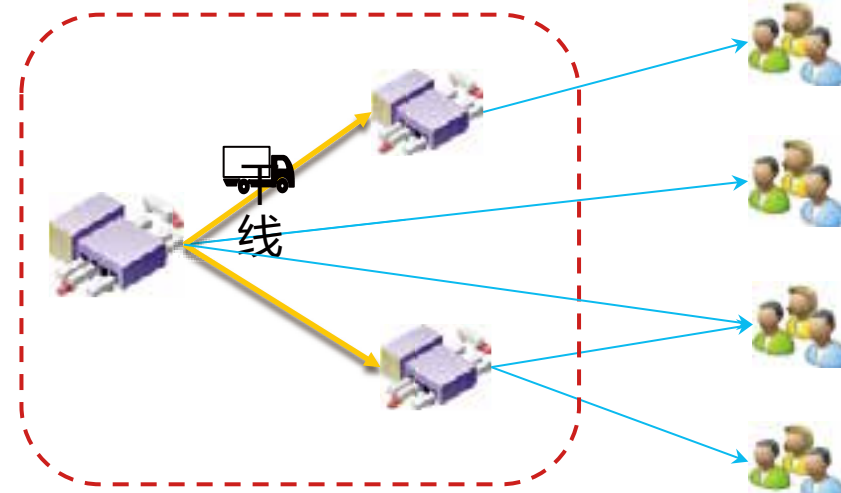
- 事件仿真
- 时间仿真
- 基于历史订单数据的

What-if scenario
simulation

智能决策是智慧物流的关键

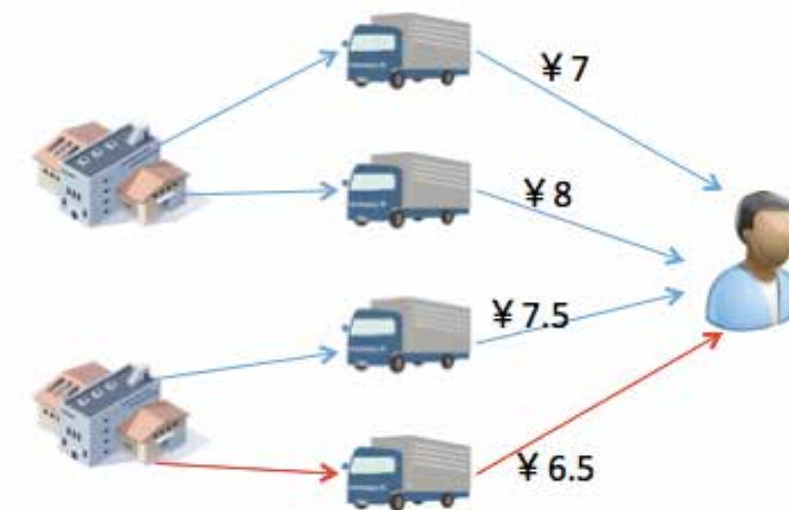
全局物流优化的基础：以数据&算法为支撑的决策引擎

库存管理



网络规划

路径规划



履行决策

机器人调度



动态路由



Decision Process Optimization
Engine

基础数据

基础算法模块

菜鸟的算法架构

Planning Problem

Strategic Planning

在哪里建仓
在哪里建分拨

Operational Planning

哪条干线要开
干线要几辆车

Execution Planning

调拨量多少
车辆怎么走

Big Data & Forecasting Problem

智能分单

智能发货

销量预测

时效预测

Multi-agent Problem

AGV调度

自动化设备调度

Real-time/Online Optimization

履行优化

箱型推荐

拣选路径优化

小件员调度

NLP & Vision

地址库

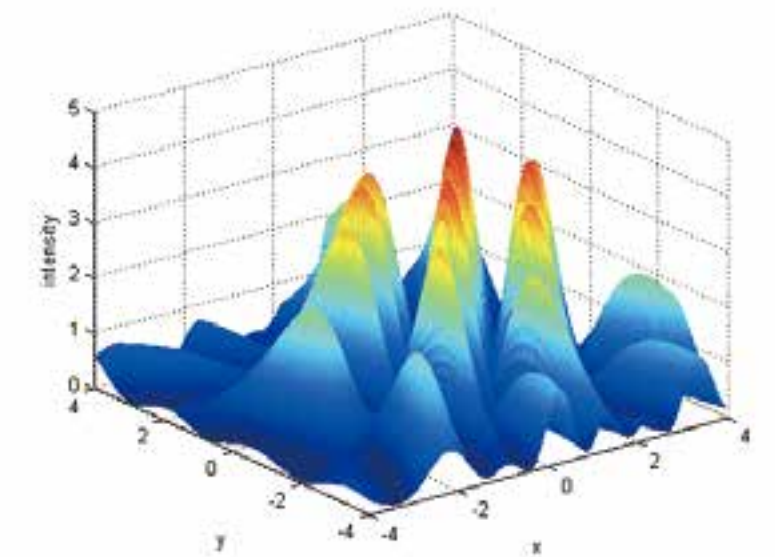
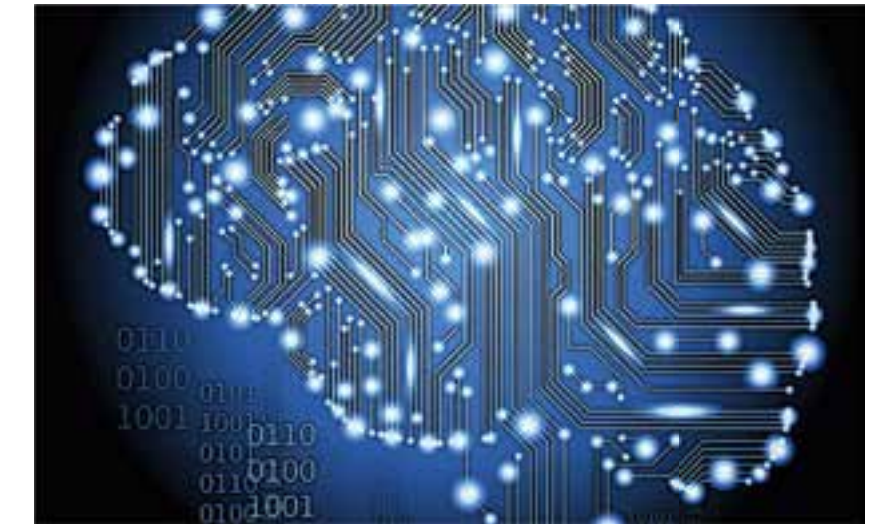
视频监控

物体感知

运动规划

物流优化问题在大数据时代被赋予的新意义

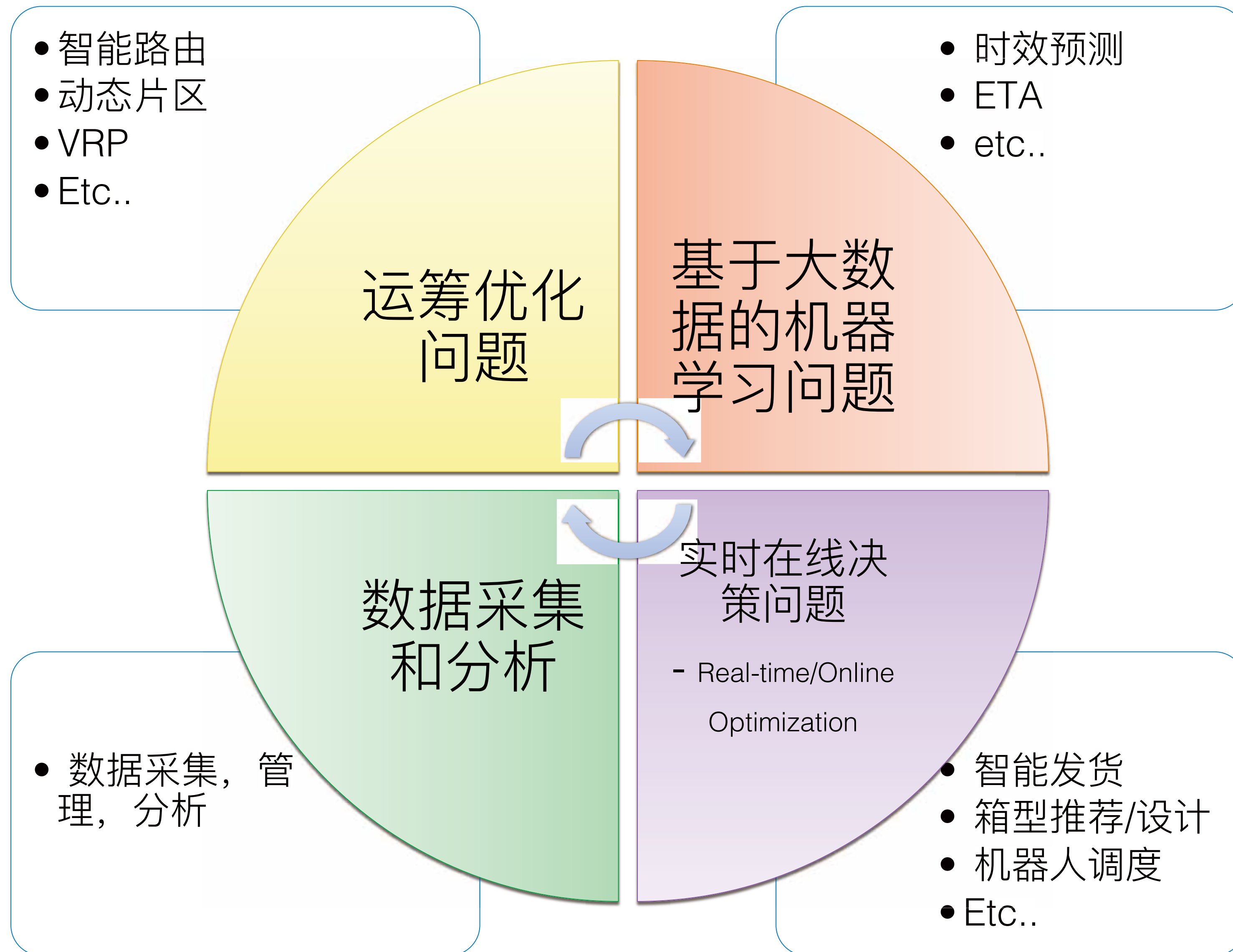
- 计算性能要求高，调用量大
- 机器学习和运筹优化算法的融合
- 利用机器学习的思想求解优化问题，从求解问题的过程中学习如何求解问题



Part 3

算法应用实例

菜鸟算法应用案例



时效预测

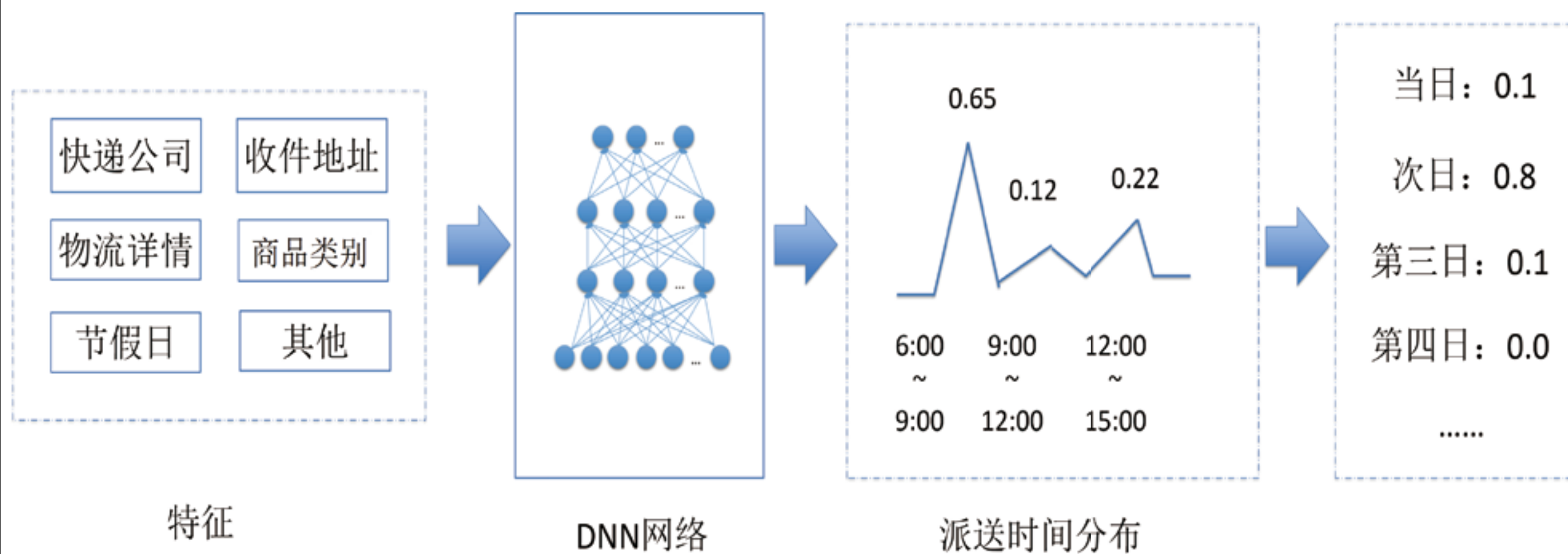
我的包裹什么时候能到？



时效预测

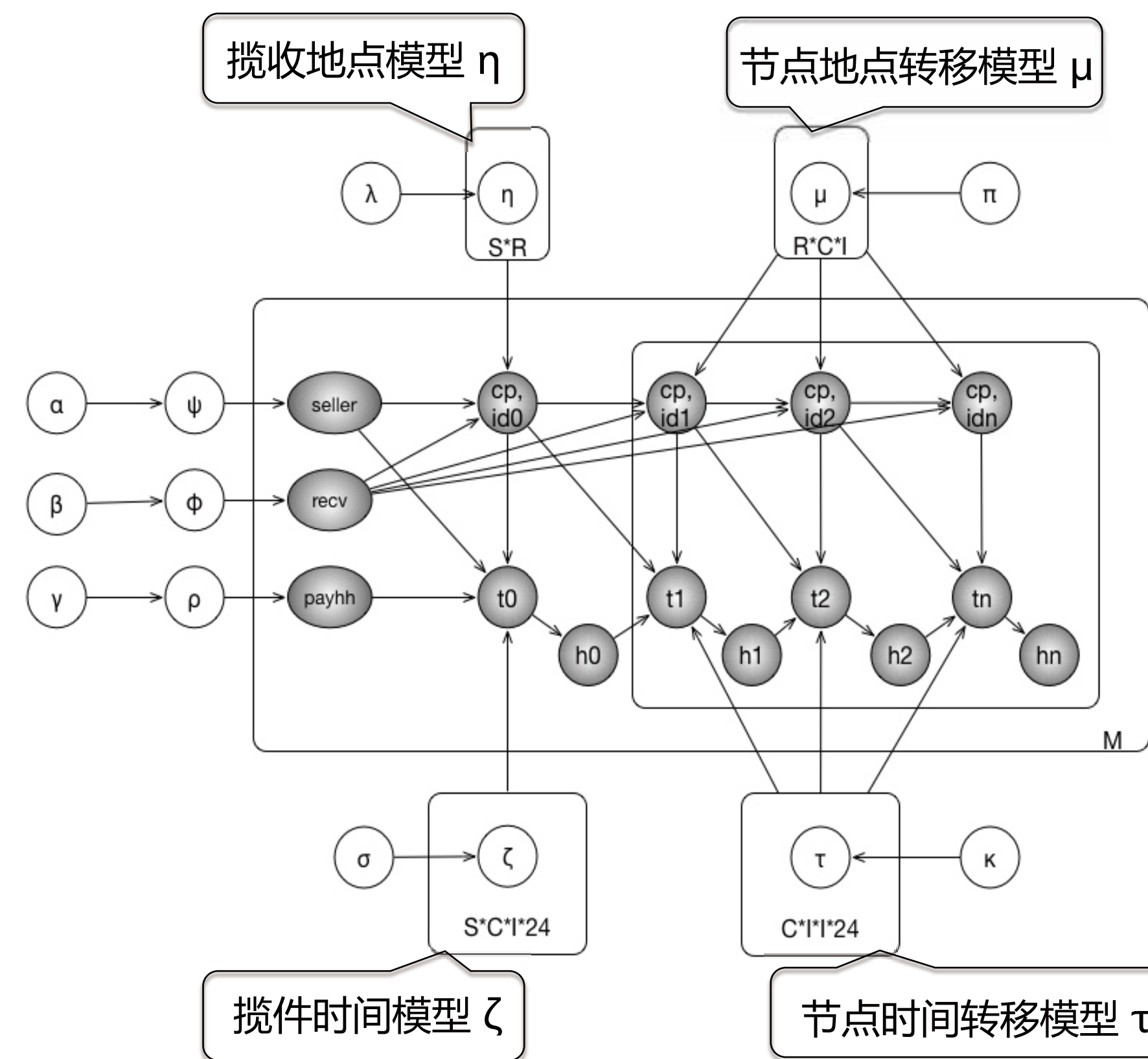
黑盒模型

白盒模型



采用了DNN多分类模型进行时效预测，以包裹当前所在网点、收货地址、快递公司、商品类别等信息作为特征输入，预计包裹派送时间，以分布的形式给出

DNN 缺少可解释性，一旦预测结果出现问题，用户还需要知道原因



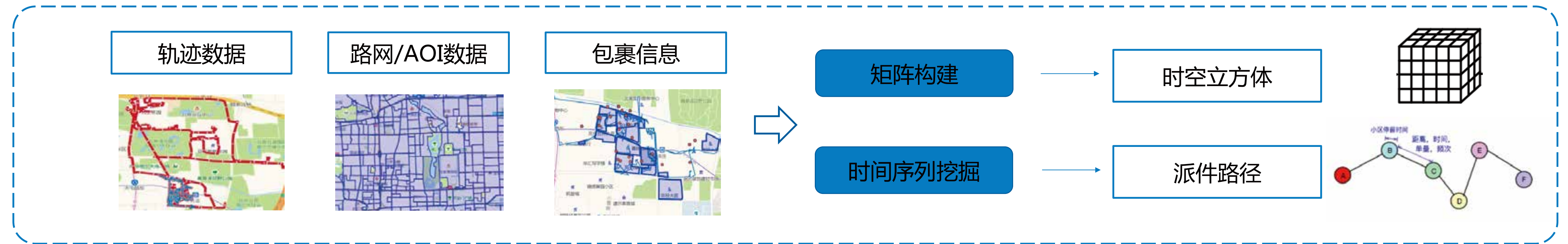
构造生成模型，模拟生订单的所有物流事件。物流运作规律基本符合一阶马尔科夫链，只看包裹现在什么状态以及要去哪，不关心包裹的过去。在订单产生后，其行为由四个模型产生：【 η, μ, ζ, τ 】

Dirichlet 分布作为模型假设，根据历史日志MLE估计模型参数

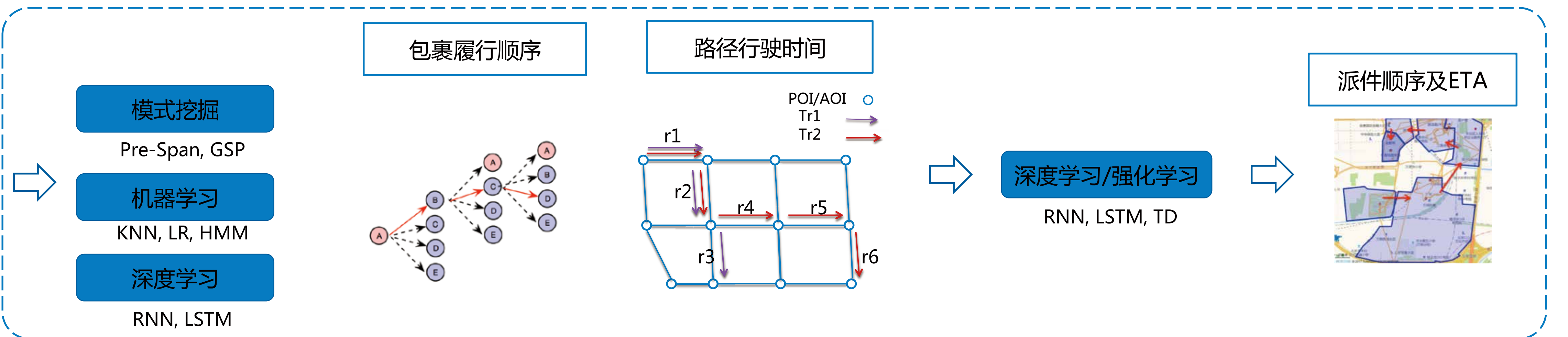
包裹送达时间预测 – 最后一公里

利用机器学习方法，让消费者随时感知包裹位置，并对包裹送达时间(ETA, Estimate Time of Arrival)进行动态预测

算法：时空数据挖掘 + 深度学习



数据预处理/特征提取



模型构建及预测

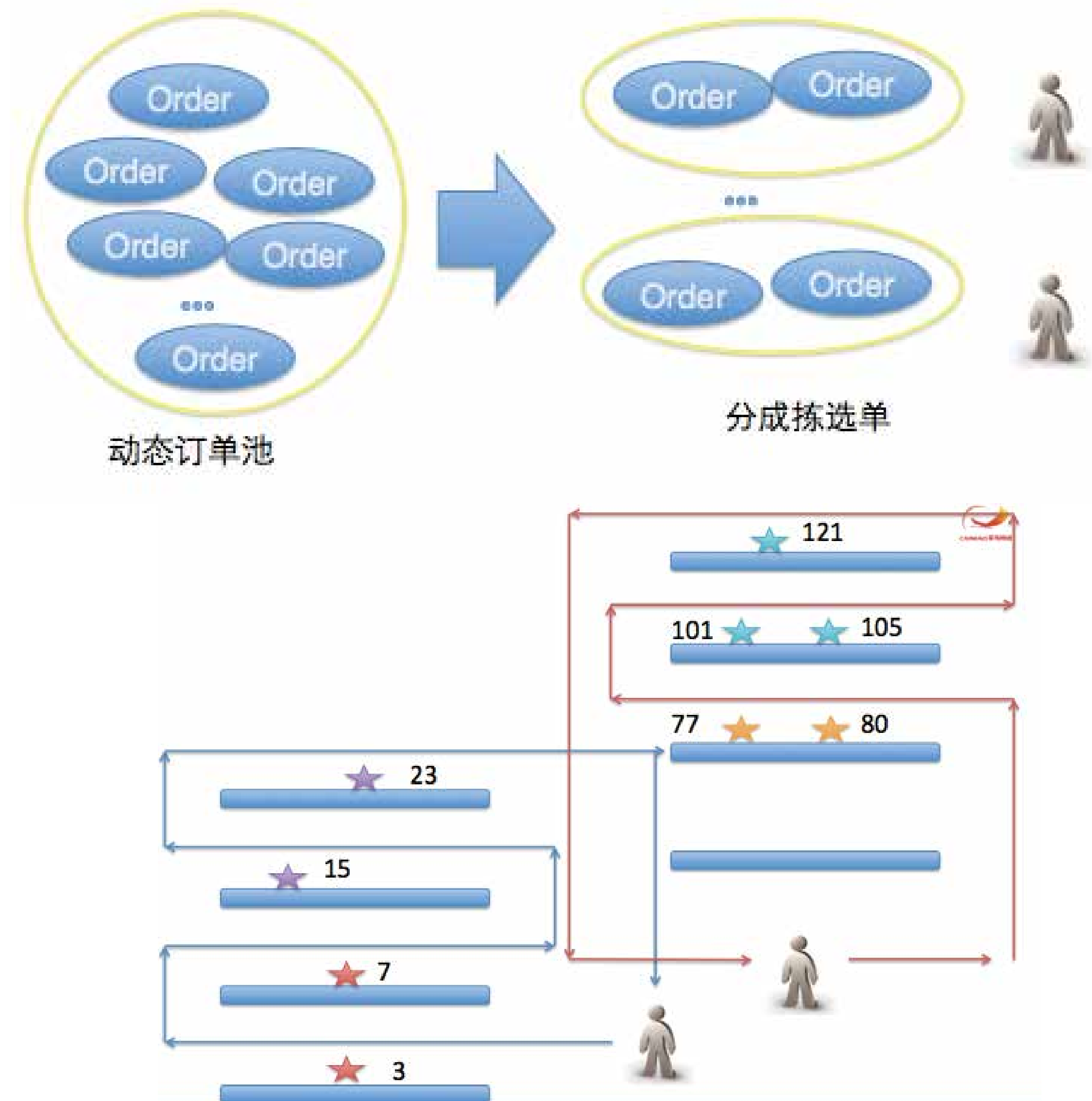
拣货路径优化

我的订单里的商品怎么找到的？



拣货路径优化

- Input : 仓库内所有未完成订单拣选任务
- Output : 哪些订单合成同一个拣选单
- 目标 : 最小化行走距离、巷道数、动碰次数等
- 动态问题 : 订单不停下发
- 每一个要拣的SKU可能在几个不同的货架上有
- Generalized Vehicle Routing Problem
- 求解方式 :
 - VRP Engine
 - Clustering + local search

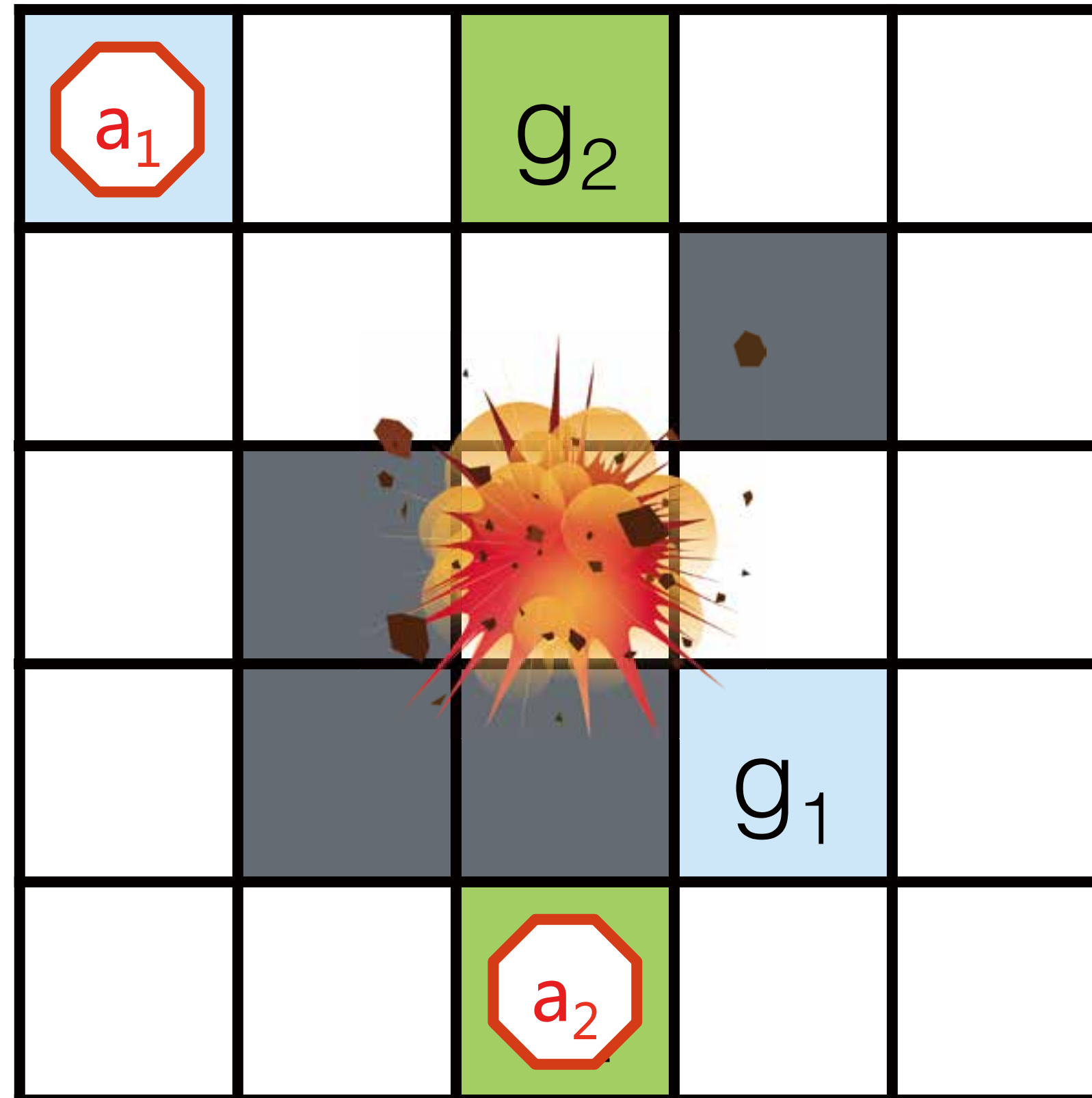


Multi-agent Path Finding (MAPF)

多智能体路径规划问题

1. 有一组机器人
2. 为每个机器人规划一条从到的路径
3. 规划的路径互不冲突
4. 使得总体代价最低

常用：*Makespan*、*Cumulative Time Cost*(累计时间)



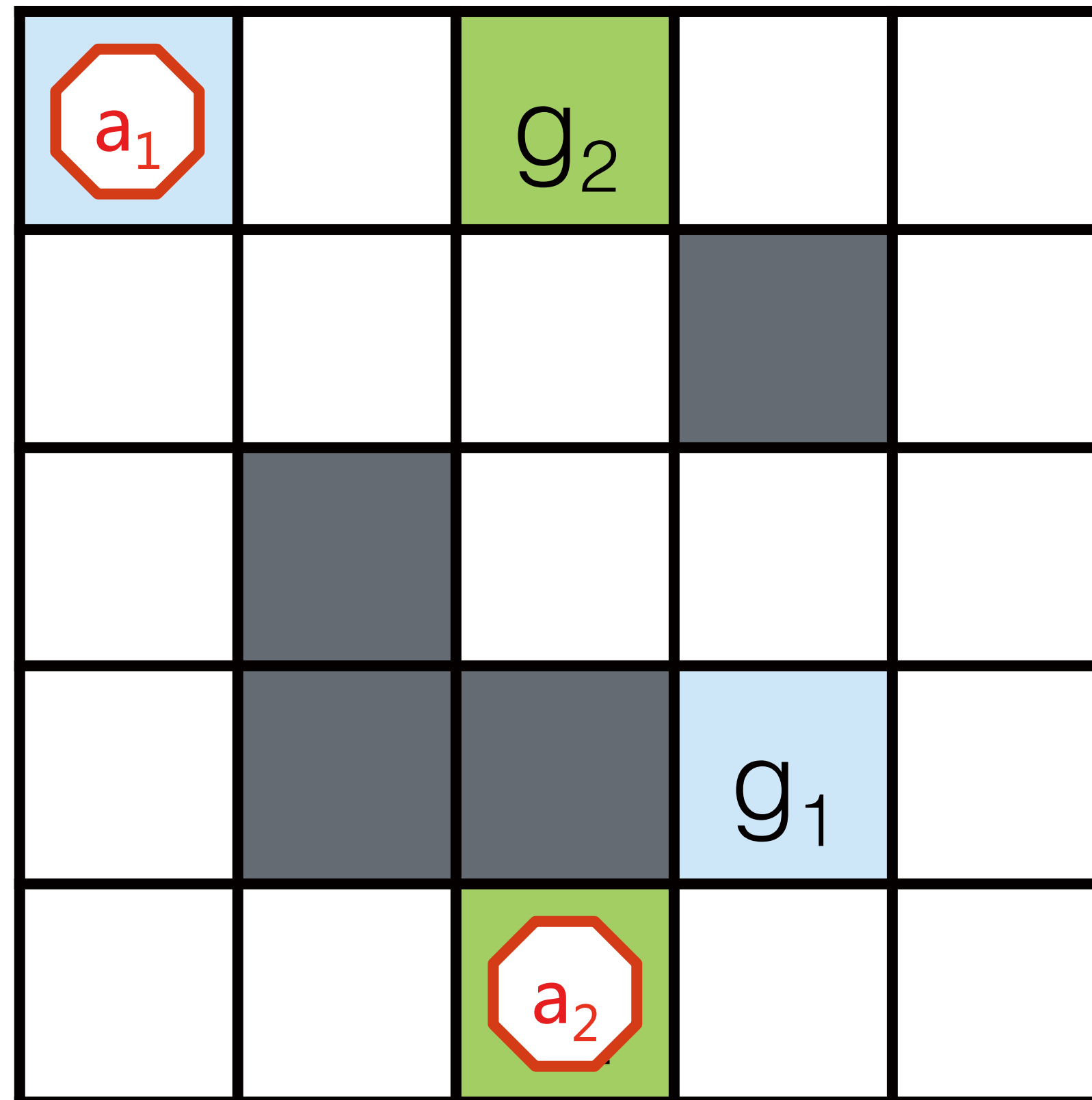
均走最短路径

Multi-agent Path Finding (MAPF)

多智能体路径规划问题

1. 有一组机器人
2. 为每个机器人规划一条从到的路径
3. 规划的路径互不冲突
4. 使得总体代价最低

常用 : *Makespan*、
Cumulative Time
Cost(累计时间)



最优解
Makespan: 8
累计时间: 14

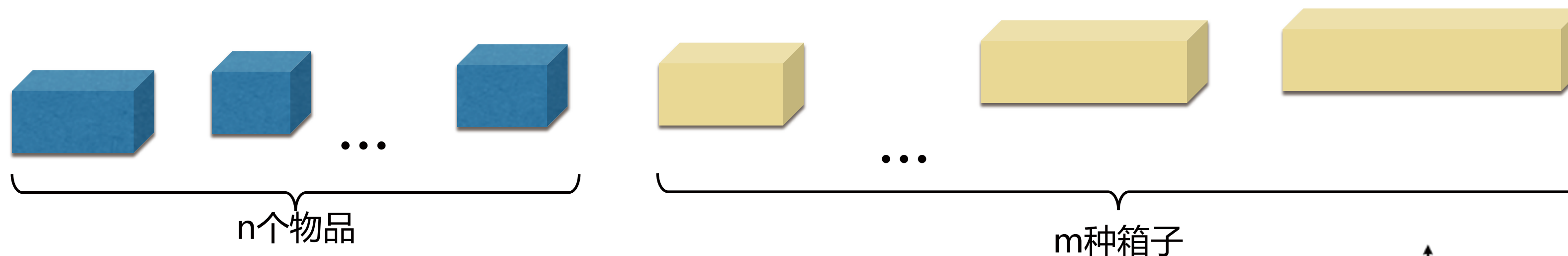
箱型推荐

我的订单里是怎么打包的？

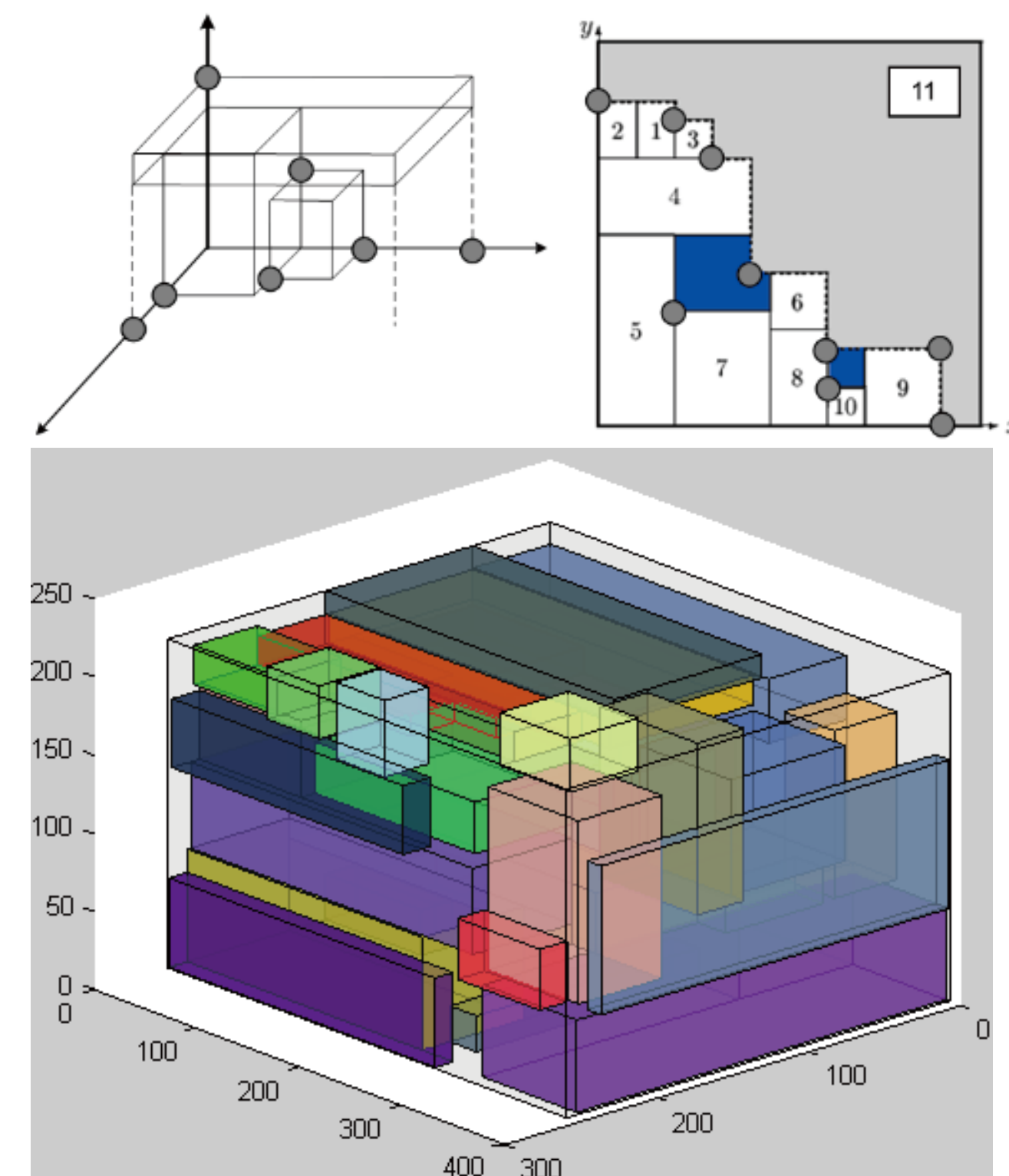


三维装箱问题

- 背景：一个订单中有 n 个物品需要装箱,有 m 种箱子可供选择

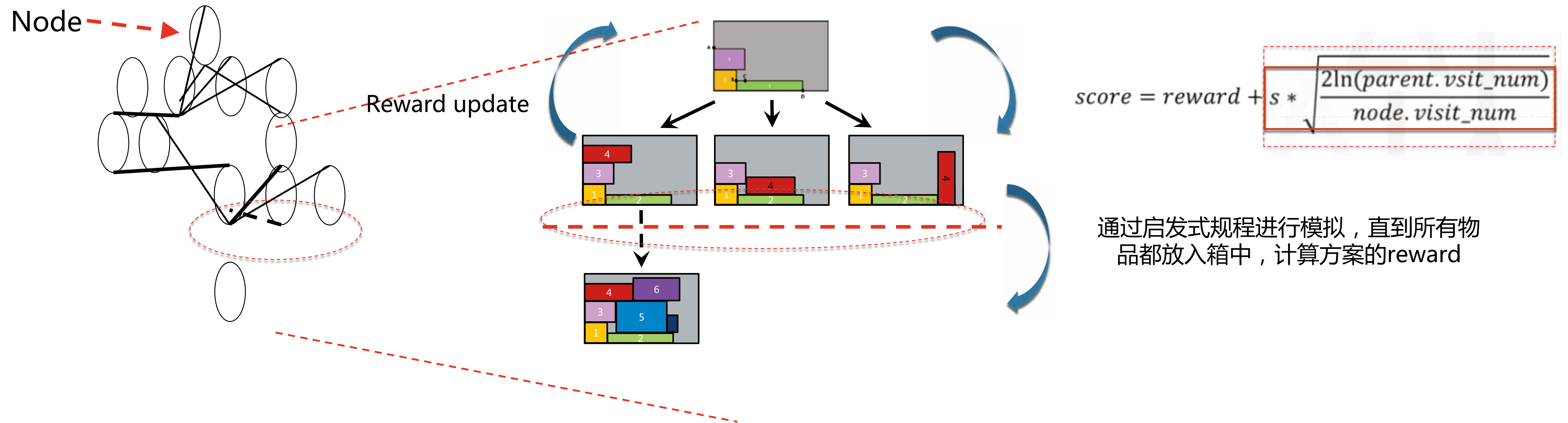


- 目标：最小化箱子花费与箱子个数
- 决策变量：摆放顺序，位置，朝向
- 约束：体积约束与质量约束
- 传统解法：
 - 构造式算法 (maxRectangle , extremePoint, etc..)
 - 元启发算法 (PSO/ACO/Genetic)



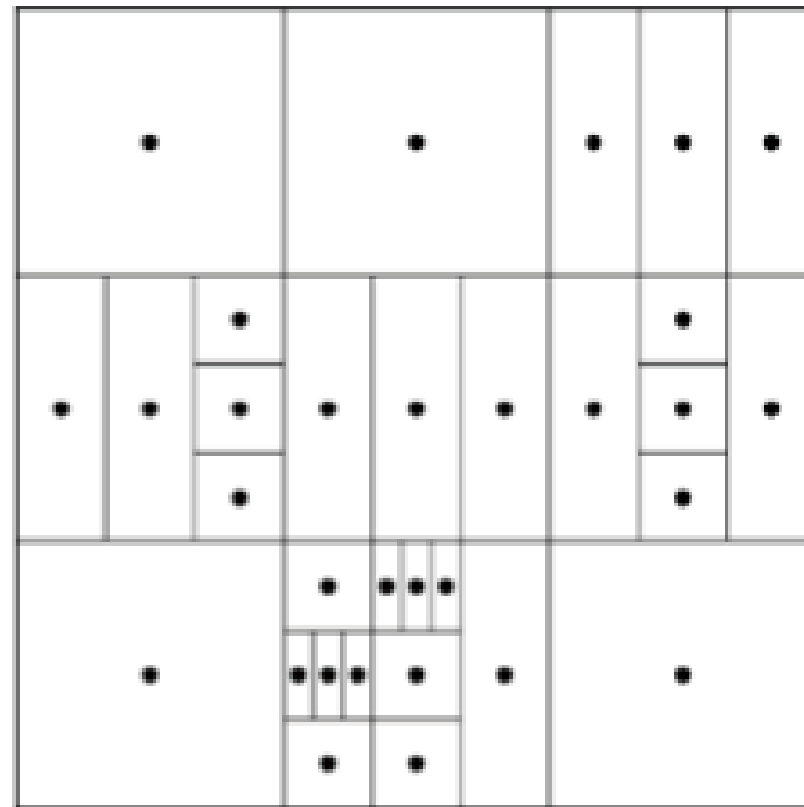
三维装箱问题

- 蒙特卡洛树搜索
- 搜索过程对应的解空间结构可以用树形图刻画



- Node记录的信息：state , parent , child , visit_num 和 reward

箱型设计



新的问题：

根据历史的订单数据，设计每个仓库最合适的箱型
优化总体包材和配送成本

目标:

最小化长时间总体成本

模型组成:

3D BinPacking Problem

包材尺寸从约束变为优化决策变量

算法应用

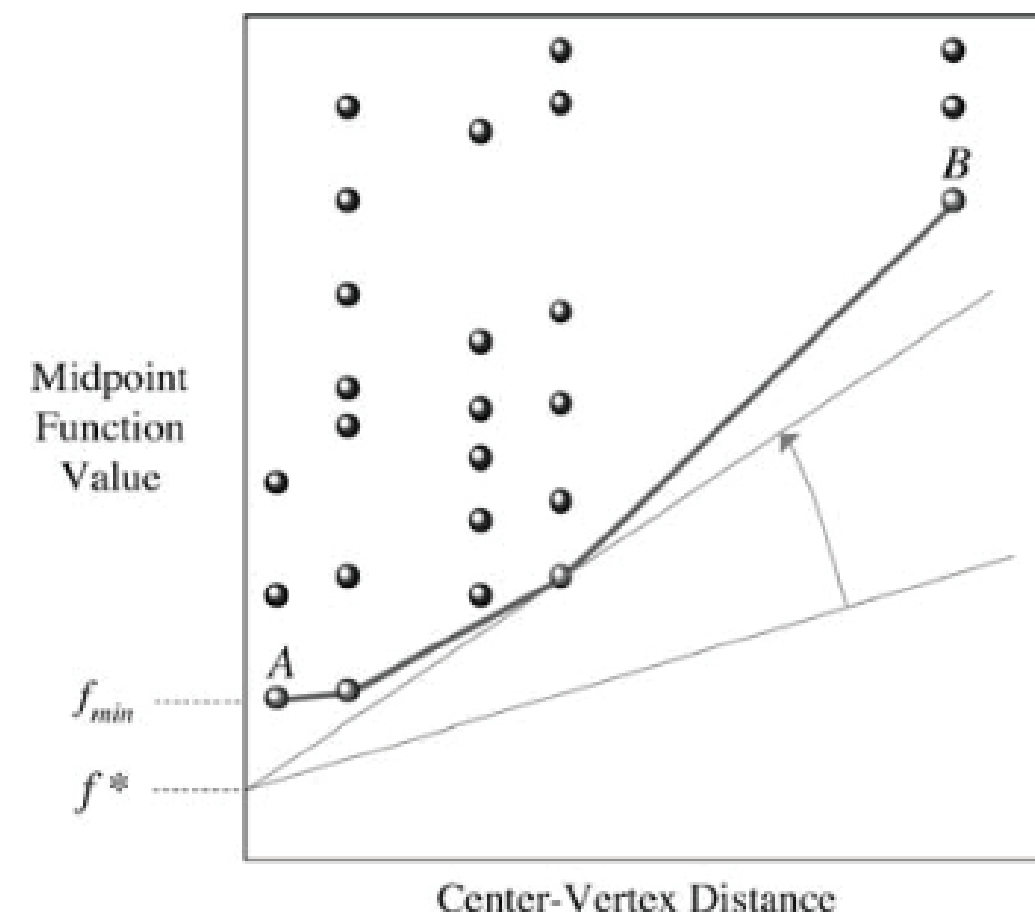
Derivative Free Optimization

Dynamic Programming + convex hull trick

$$\text{cost}(B_j, j) = \min\{\text{cost}(B_{j-1}, j-1) + \text{area}(B_j) * (\text{sum}(B_j) - \text{sum}(B_{j-1}))\}$$

时间复杂度

$$O(K * L^2 * W^2 * H^2), \text{ 可优化到 } O(K * L * W * H * \log(L) * \log(W) * \log(H))$$



基于数据的个性化包材设计

车辆路径规划

我的订单里是怎么运输的？



车辆路径规划

Vehicle Routing Problem

输入参数

需要服务的顾客集合、可提供车辆数、
 车辆容量、固定成本等

输出结果

车辆具体行走路线、无法服务的顾客集合

优化目标

最小化费用、车辆数、距离或者时长。

约束条件

货车载重、容积、数量
 货车最大行驶距离
 最早、最晚送货或取货时间

CVRP

Capacitated VRP

VRPTW

VRP with Time Windows

VRPPD

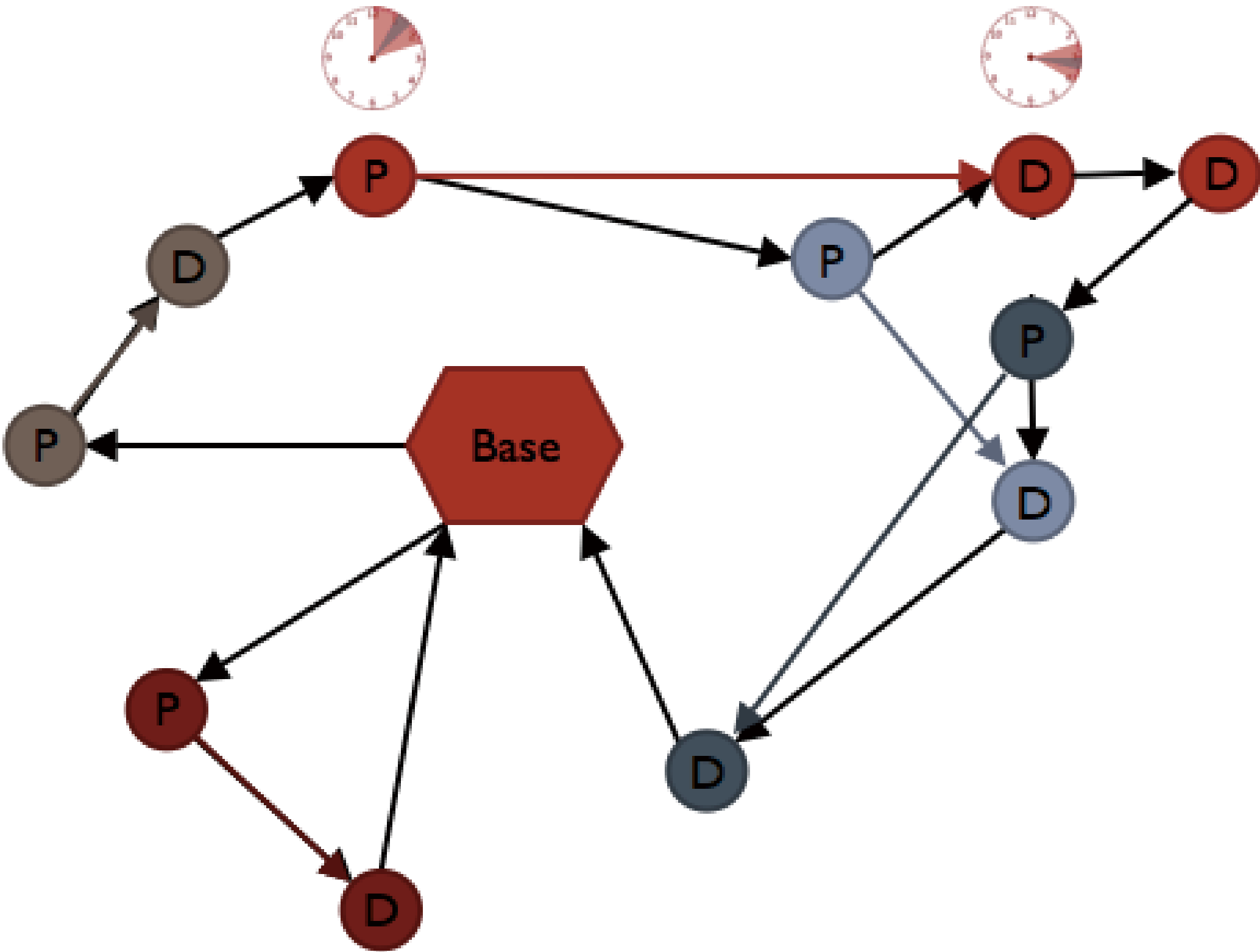
VRP with Pickup and Delivery

VRPPDTW

Time Windows + Pickup and Delivery

MDVRP

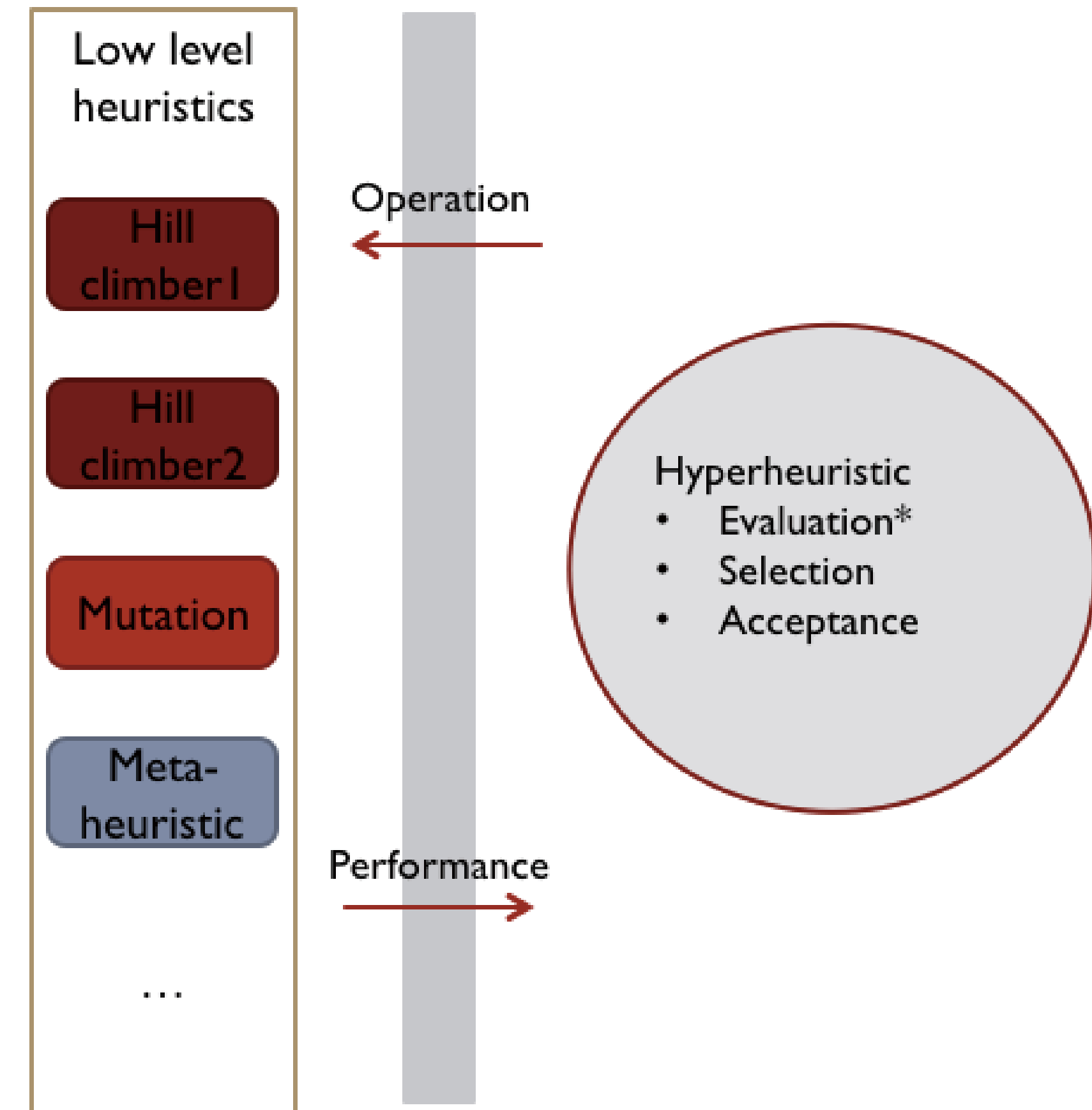
Multi-Depot VRP



车辆路径规划

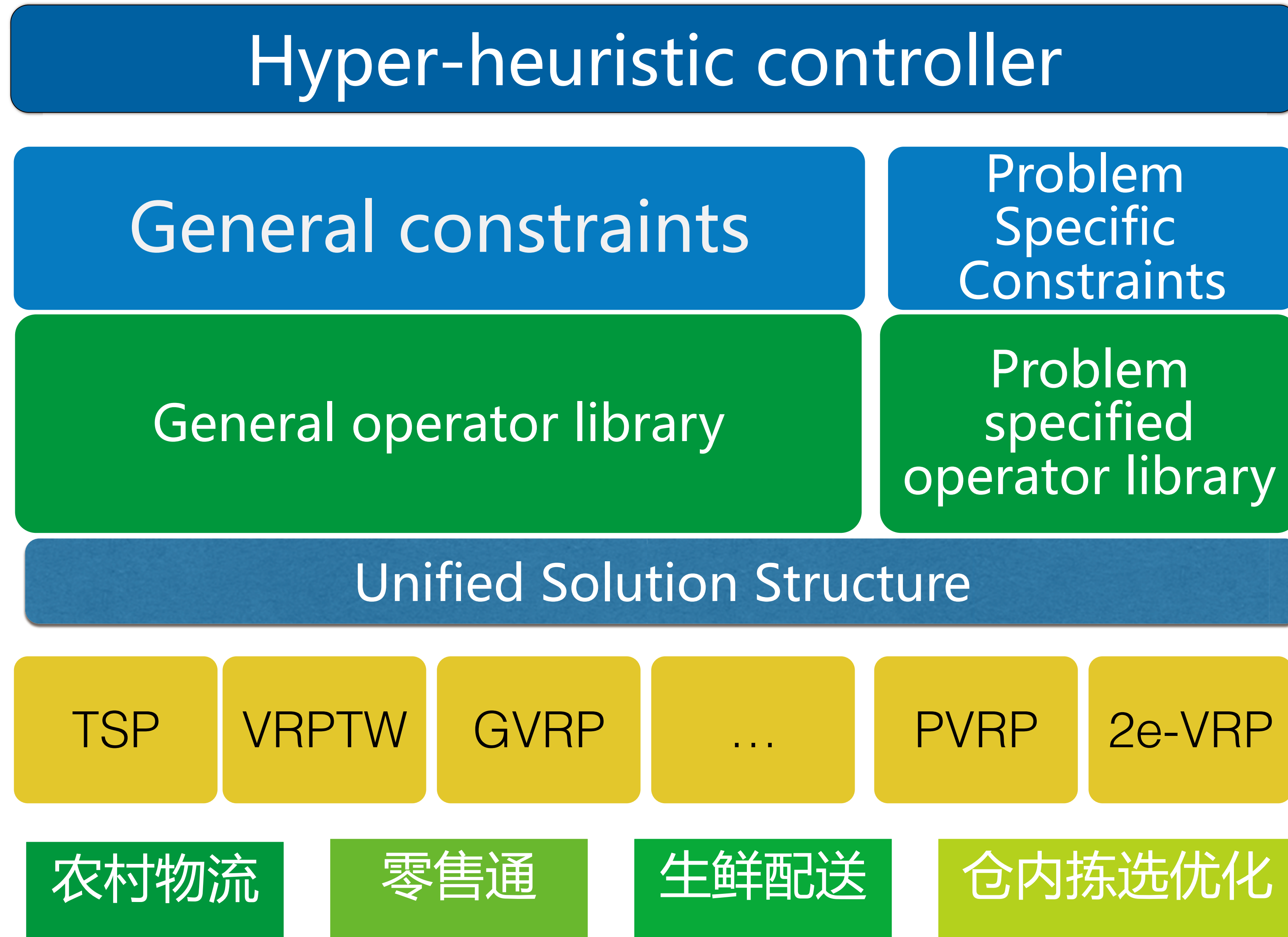
Hyper-Heuristic Algorithm

- **Some** (meta)heuristics perform better for some type of problems.
 - For the same problem, **different (meta)heuristics** perform better for **different instances**.
 - **Different (meta)heuristics** perform better at **different stages** of a search for the same instance.
-
- Initial Solution
 - Iterative Improvement
 - Use some low level heuristic (LLH) operators to find a neighborhood, based on the score of those operators
 - Apply LLH, evaluate its performance
 - Accept/Reject
 - Update scores for LLH
 - Termination



车辆路径规划

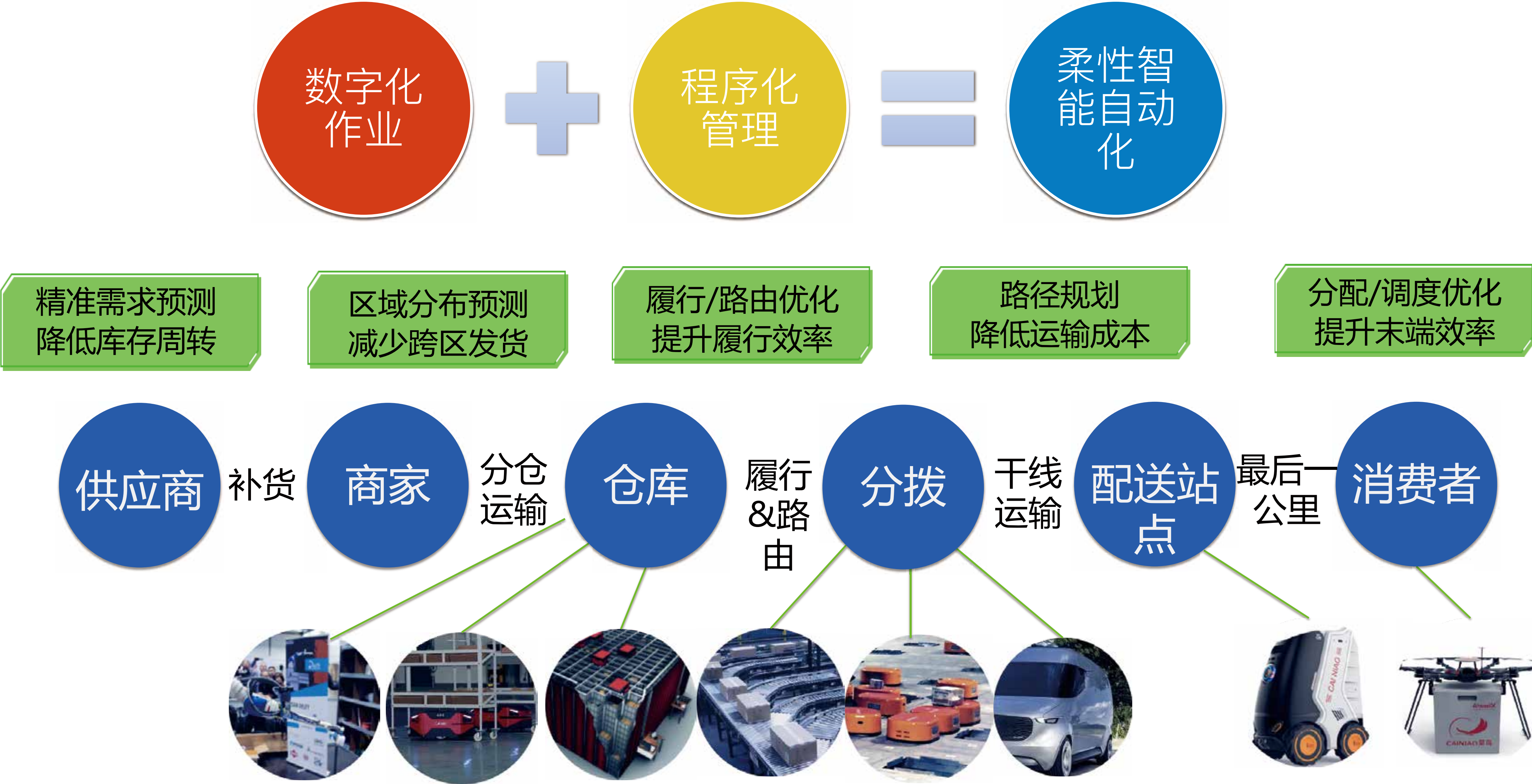
VRP算法的抽象&应用



Part 4

未来展望

未来物流大脑图景



人工智能方向 @未来物流大脑图景

1. 数据驱动 (ML) + 模型驱动 (OR) 的有机结合
2. 算法产品化问题
3. 大规模分布式优化求解器
4. 多模态多任务的机器学习问题
5. 去中心化的多智能体学习问题

运筹优化
算法

机器学习
算法

机器视
觉算法

大数据处理
引擎

NLP算
法

调度系统
开发

机器人
控制



网络通
讯

嵌入式
软硬件

系统流程
架构

机械设
计

IOT&
端计算

阿里巴巴跨境B2B中的算法技术

国际技术事业部 资深算法专家 守清

目录



阿里巴巴的
国际化战略

.....



跨境电商

.....



流量|搜索|广告|风控算法

.....



多语种NLP

.....



未来的战场

● 阿里巴巴国际化战略

● 什么是阿里巴巴的国际化（全球化）

普惠

可持续

健康&快乐

● 阿里巴巴20年的目标——第五大经济体

20亿消费者

1亿就业

1千万企业盈利

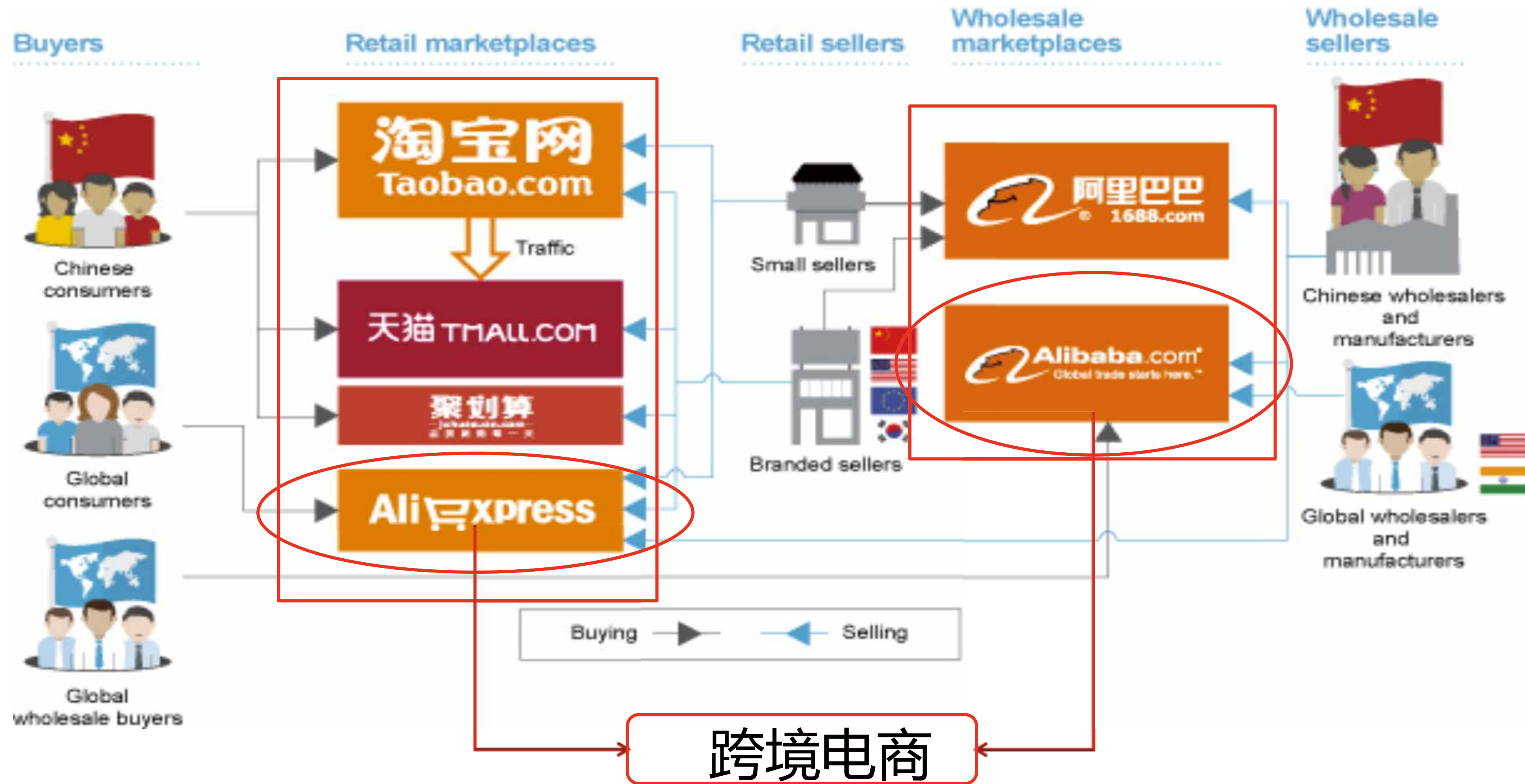
● 国际化战略中的跨境电商：全球买，全球卖

eWTP

目录

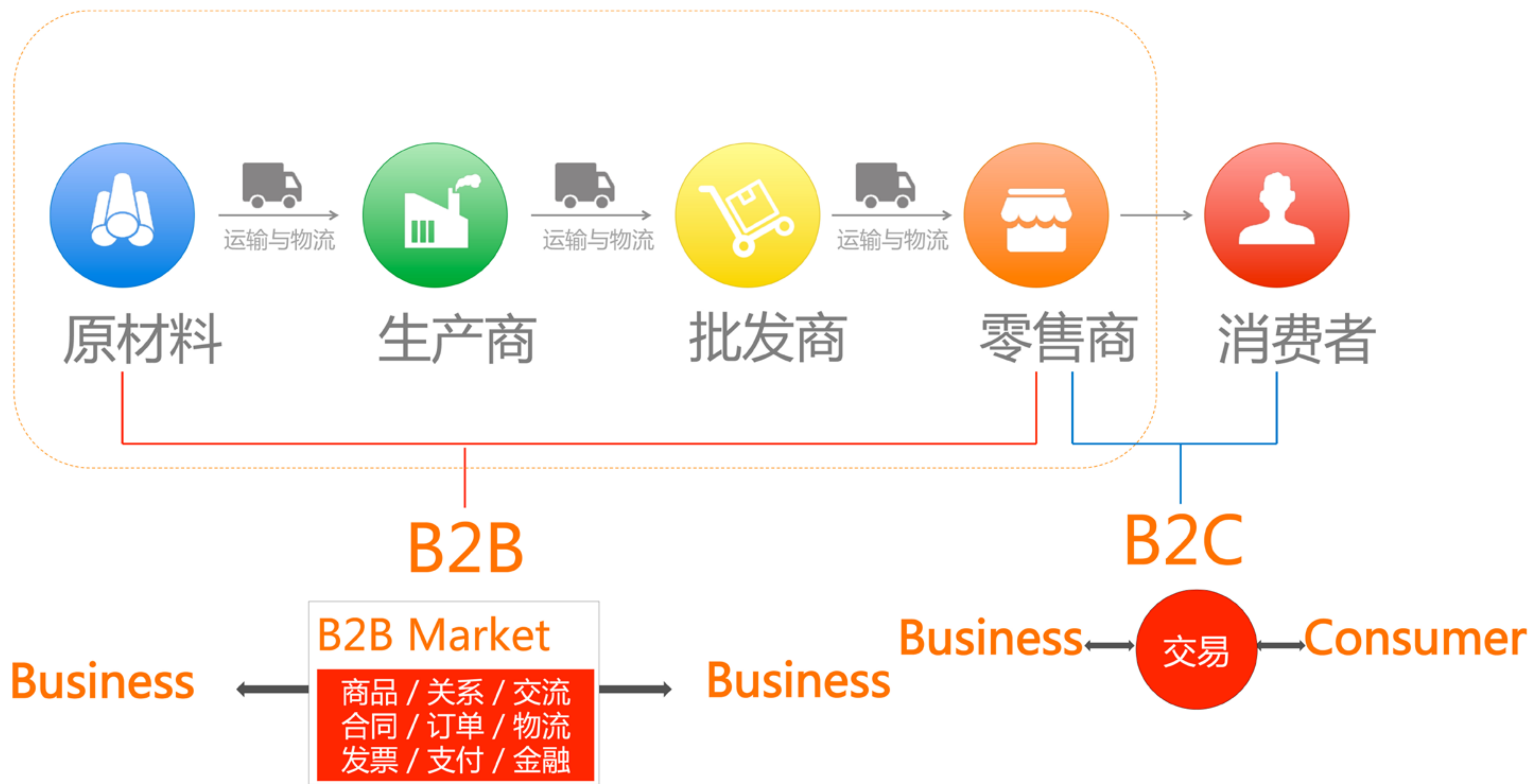


跨境电商的业务位置



让天下没有难做的生意！

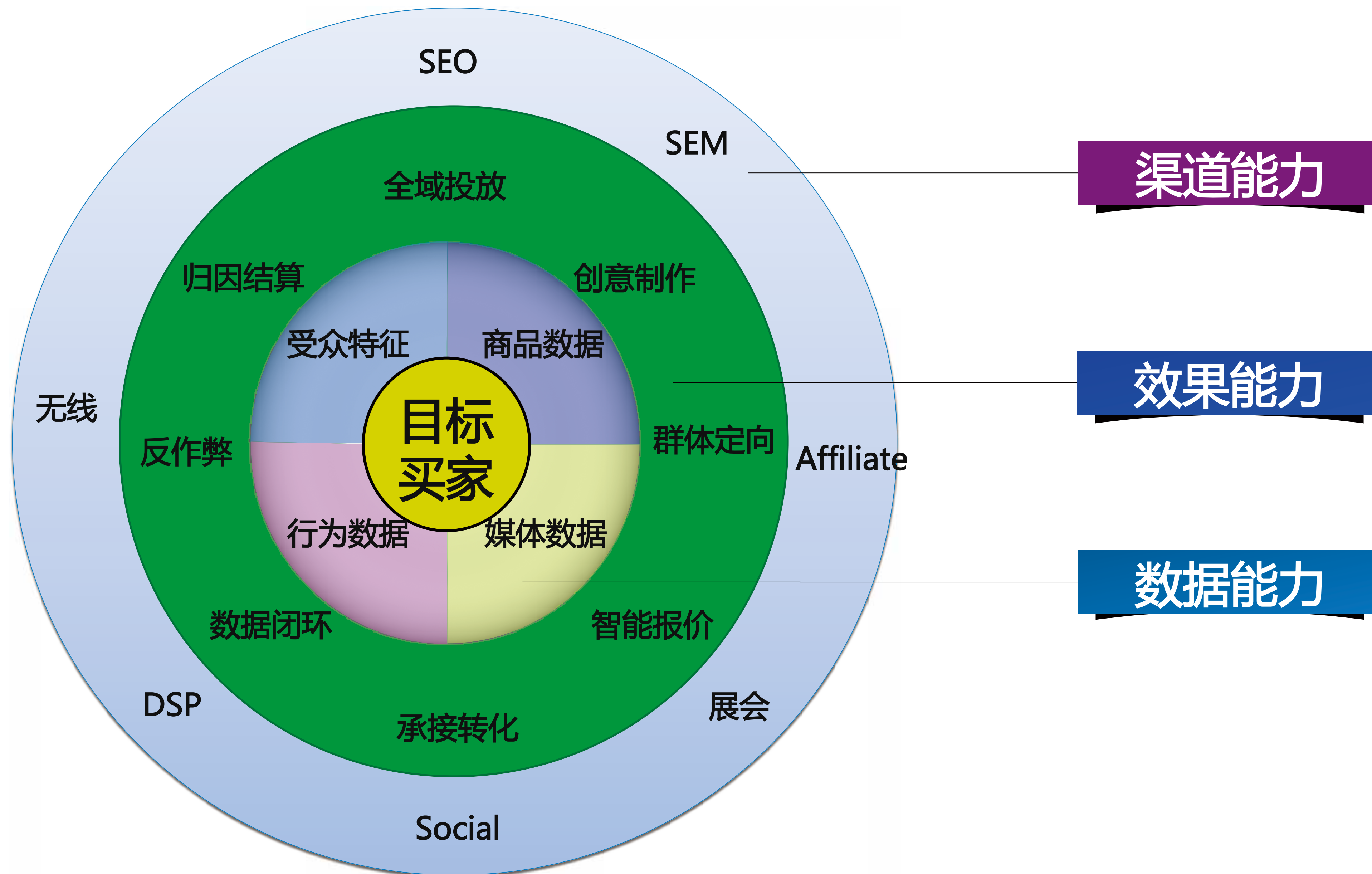
● 跨境电子商务B2B/B2C

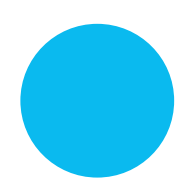


目录

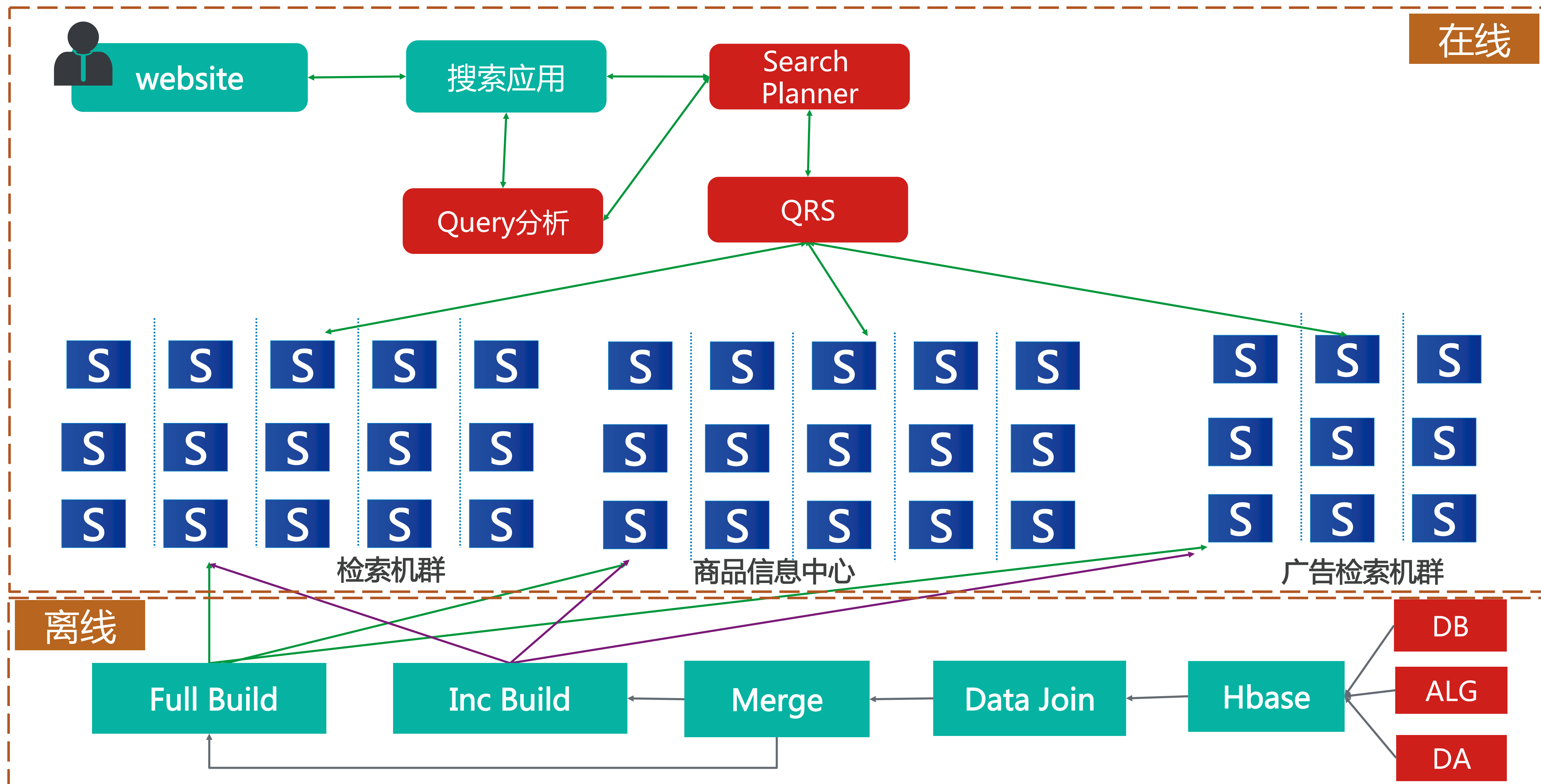


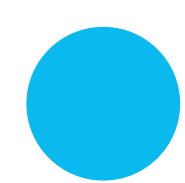
● 精准引流



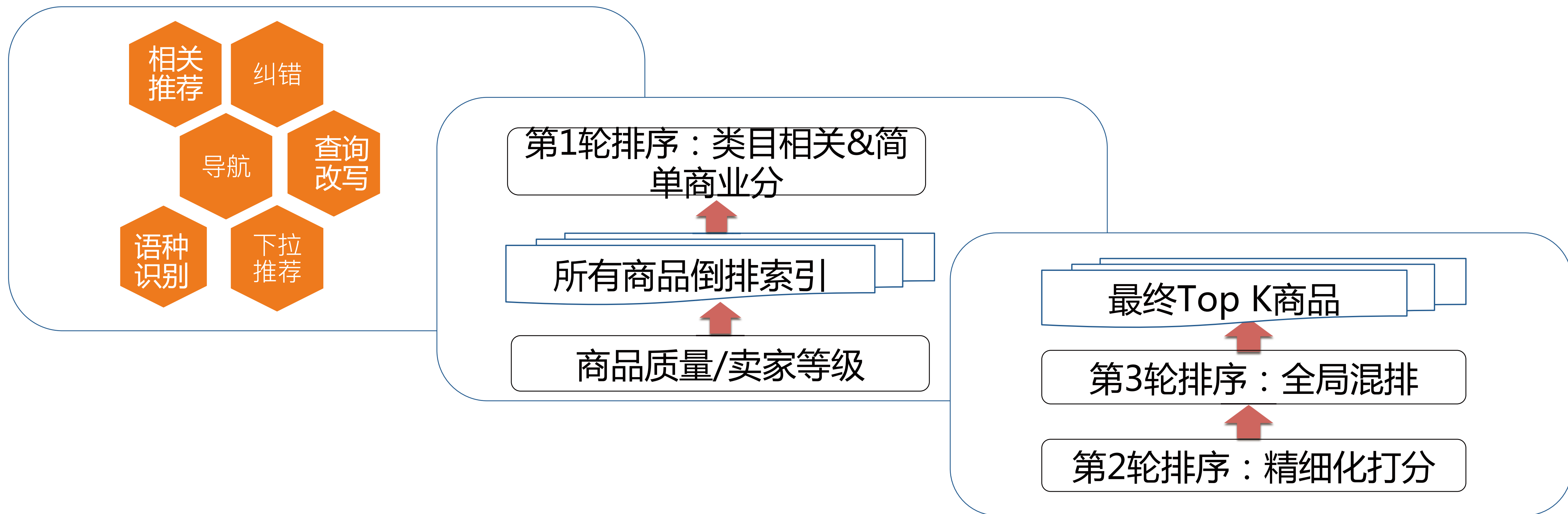


高效搜索的整体架构





高效搜索的算法核心链路

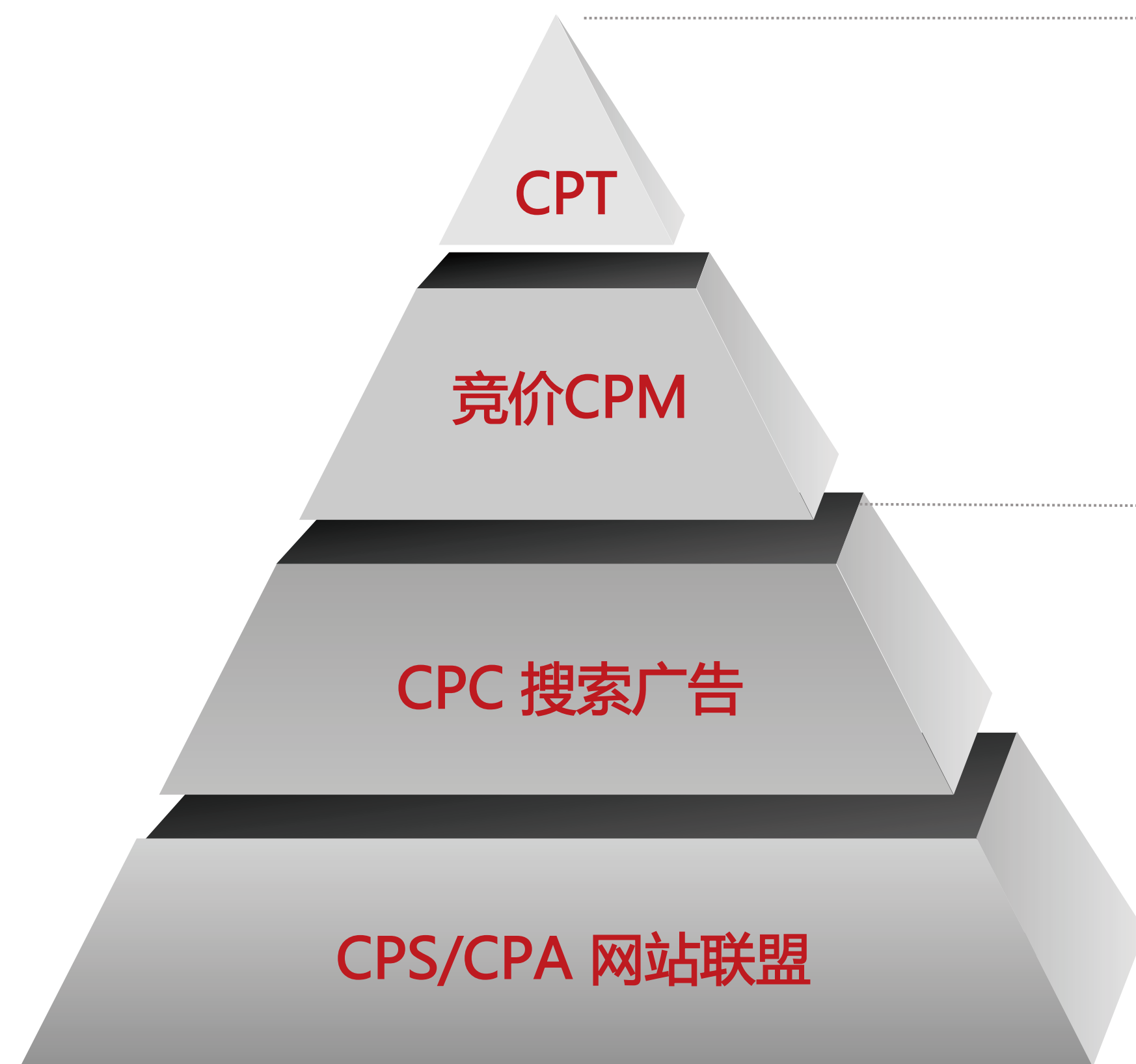


Query显性的分析，包括语种识别、纠错、导航、相关推荐等。

第1轮排序发生在单searcher上，输入商品集合较大，负责粗筛选，用最少的商品覆盖最好的结果集合。

第2轮排序发生在单个searcher上，输入商品集合中等，负责精筛选，追求准确排序。

第3轮排序发生在merger，输入商品集合较小，是各来源商品的汇总，追求全局质量和总体满意度。



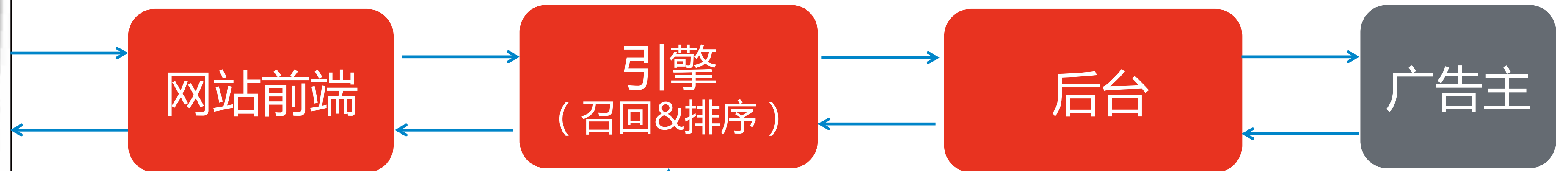
品牌广告

全网标王
线下广告
全网品牌营销
顶级展位

效果广告

搜索推广
联盟广告
整合营销
一键推广

● CPC效果广告模型



$$RankingScore = eCTR^T * bidprice$$

- 反馈信号位置的bias
- 线性模型&非线性模型
- 特征的筛选和组合
- 千人千面的个性化
- 大规模数据的处理能力
- 冷启动

- 合理分配广告主预算到各时间段
- 对流量精准分析，智能出价，提升投放性价比
- T值实时动态调整各部分作用

● 商机分配策略的抽象层次



● 商机分配的算法模型结构（e.g. 线性层次模型）

分配综合分

e.g. 分配的线性层次模型

算分模型的
层次结构

干预分

$$R(\vec{x}) = \sum_{i=1}^n \sum_{j=1}^k w_{ij} x_{ij} + D(tag)$$

文本匹配

类目匹配

个性化

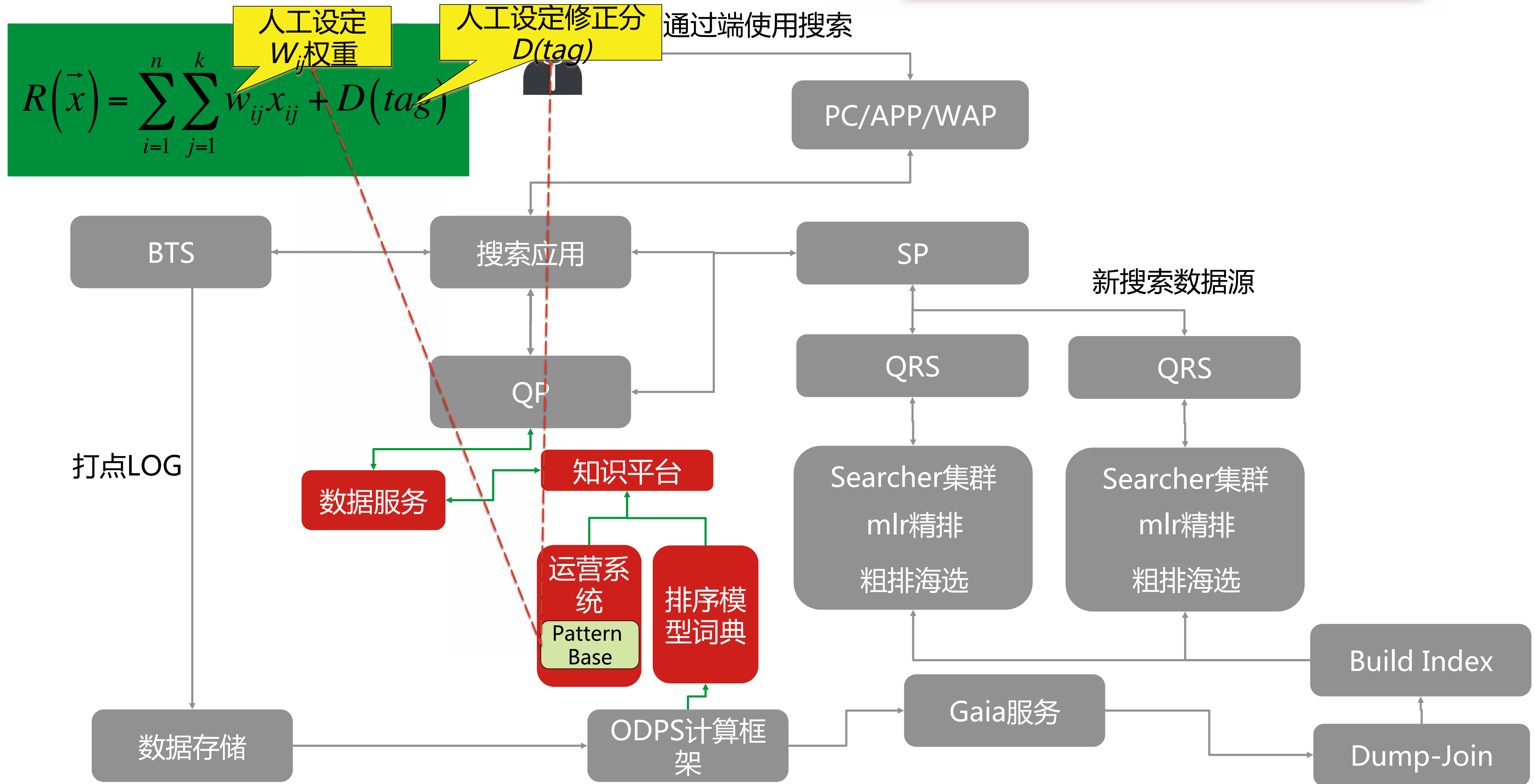
类目反馈分布数据（ $\langle q, c \rangle$ ）

文本匹配特征，商品历史点击数据，询盘/交易数据，关联卖家服务能力，纠纷，好评，卖家等级等（ $\langle q, offer \rangle ; \langle offer \rangle$ ）

用户行为路径数据（ $\langle u, offer \rangle ; \langle u, s \rangle$ ）

● 商机分配策略的能力开放

已知：人工先验知识形式化



● 商机分配策略的在线学习

半知：在线学习机制

在线学习 w_{ij} 权重

$$R(\vec{x}) = \sum_{i=1}^n \sum_{j=1}^k w_{ij} x_{ij} + D(tag)$$

数据中心

NearLine



实时点击日志



实时交易日志



实时加购/收藏日志

基于参数服务的增量训练

Δt

用户某种偏好的实时+全量融合模型

T

生成动态意图

用户分析

调用

偏好商品曝光资源控制

偏好类型商品加分

偏好类型商品分Quota

偏好显性化

搜索应用

Search Planner

搜索结果汇聚

排序精排

排序粗排

Online

Offline

基于参数服务的FTRL全量训练



历史点击日志



历史交易日志



历史加购收藏日志



其他日志

● 商机分配策略的E&E机制

未知：基于特征选择、模型选择的E&E

不断试探模型来寻优

$$R(\vec{x}) = \sum_{i=1}^n \sum_{j=1}^k w_{ij} x_{ij} + L$$

模板信息

场景信息

条件信息

动作信息

意图模板，动作信息是某个策略

$$R(\vec{x}) = g_0$$

共轭模板簇

流量试探模板能力

流量试探模板能力

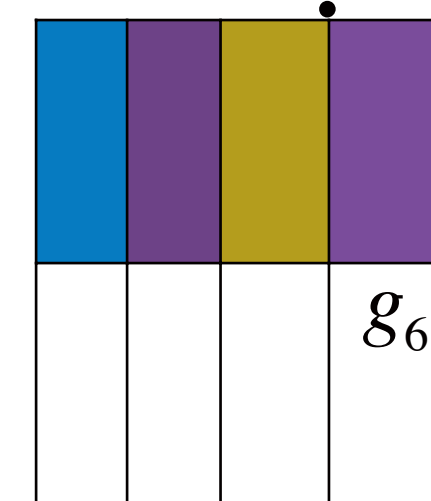
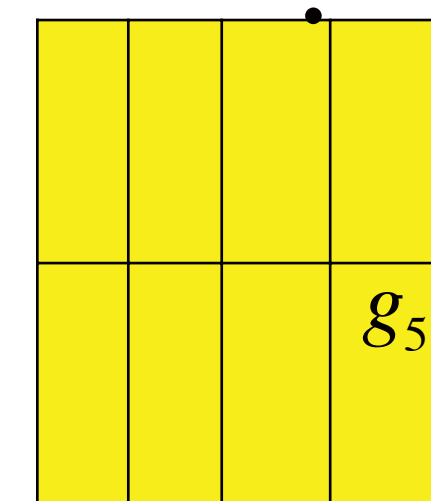
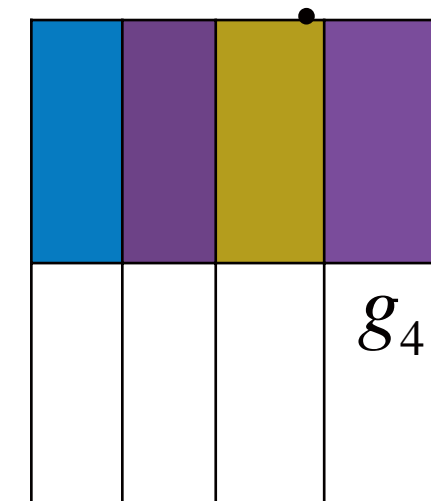
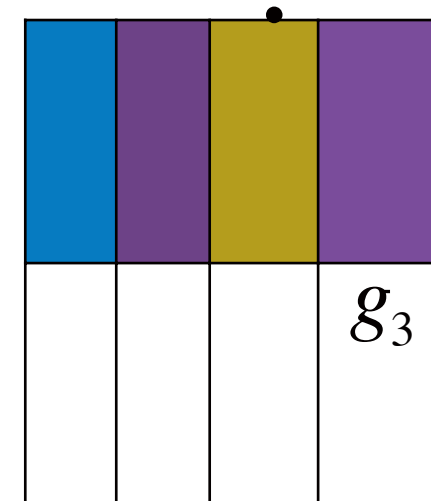
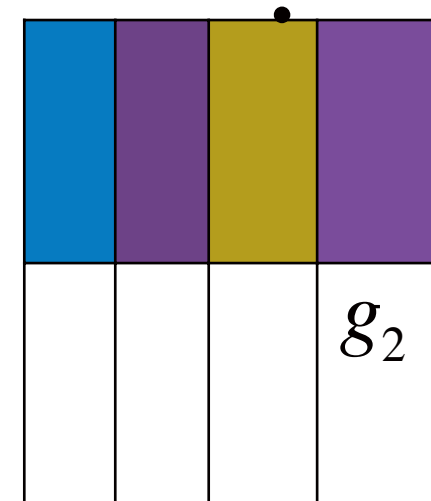
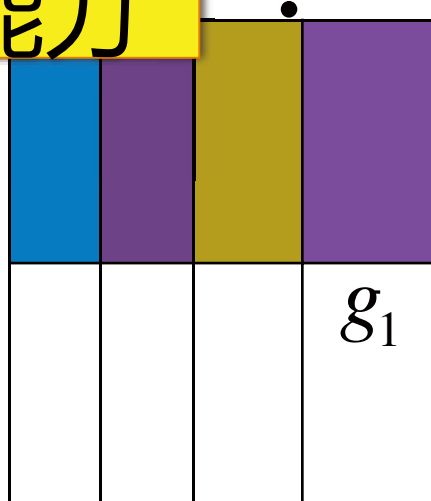
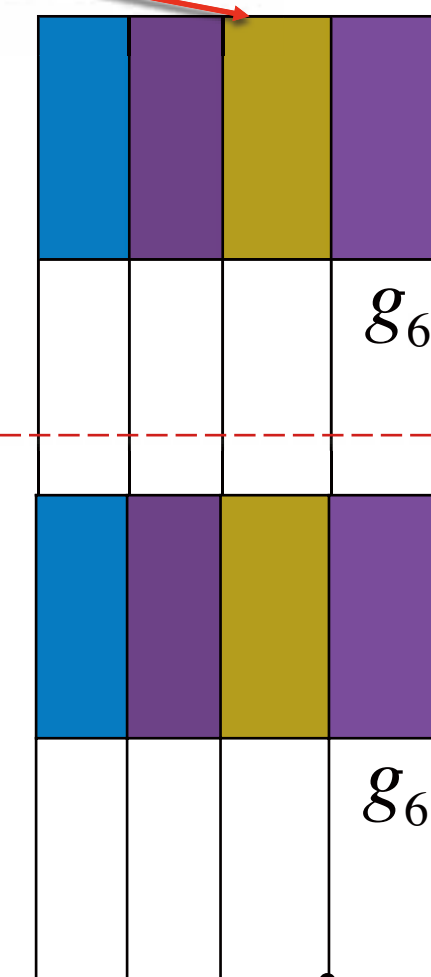
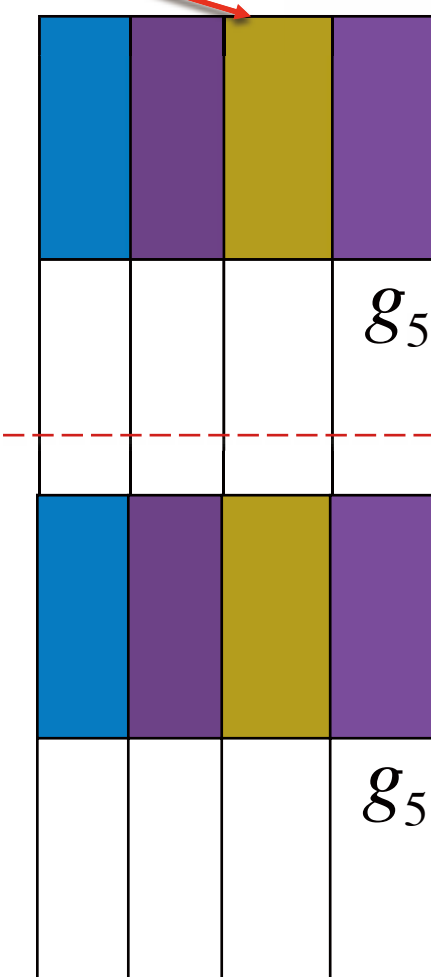
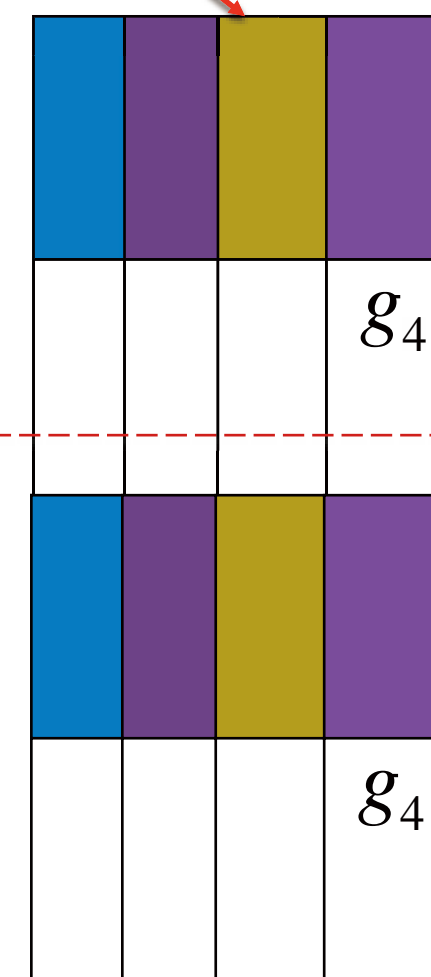
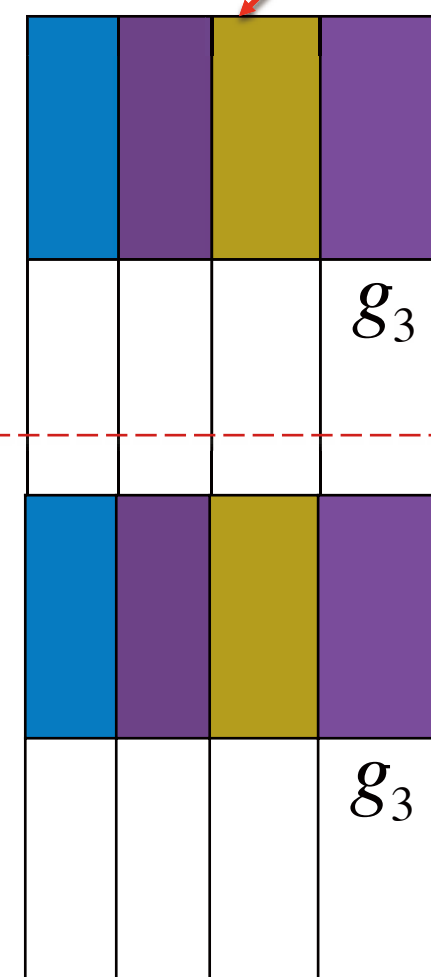
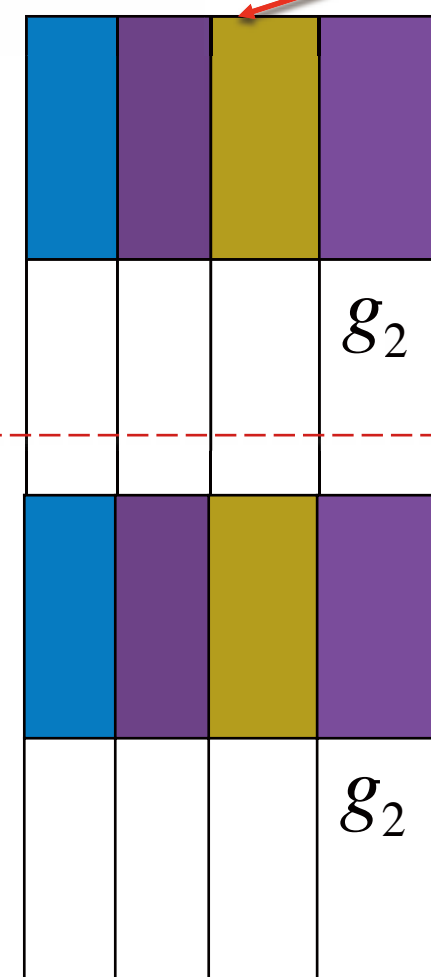
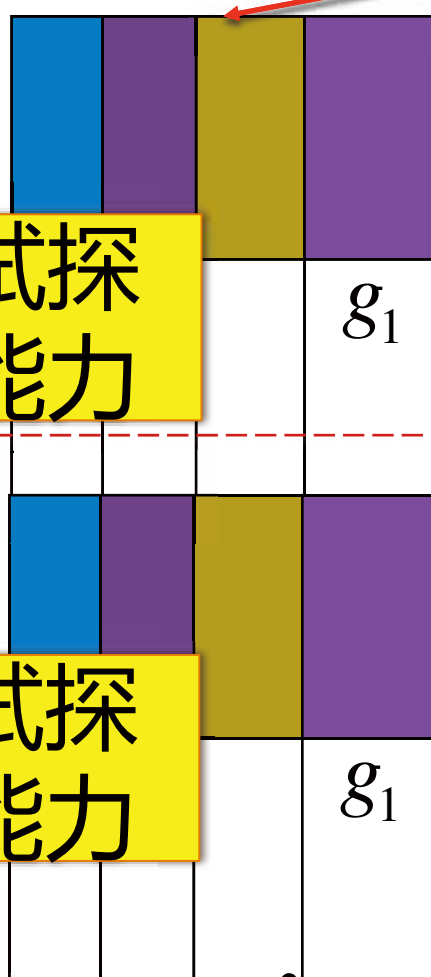
流量试探模板能力

T_0

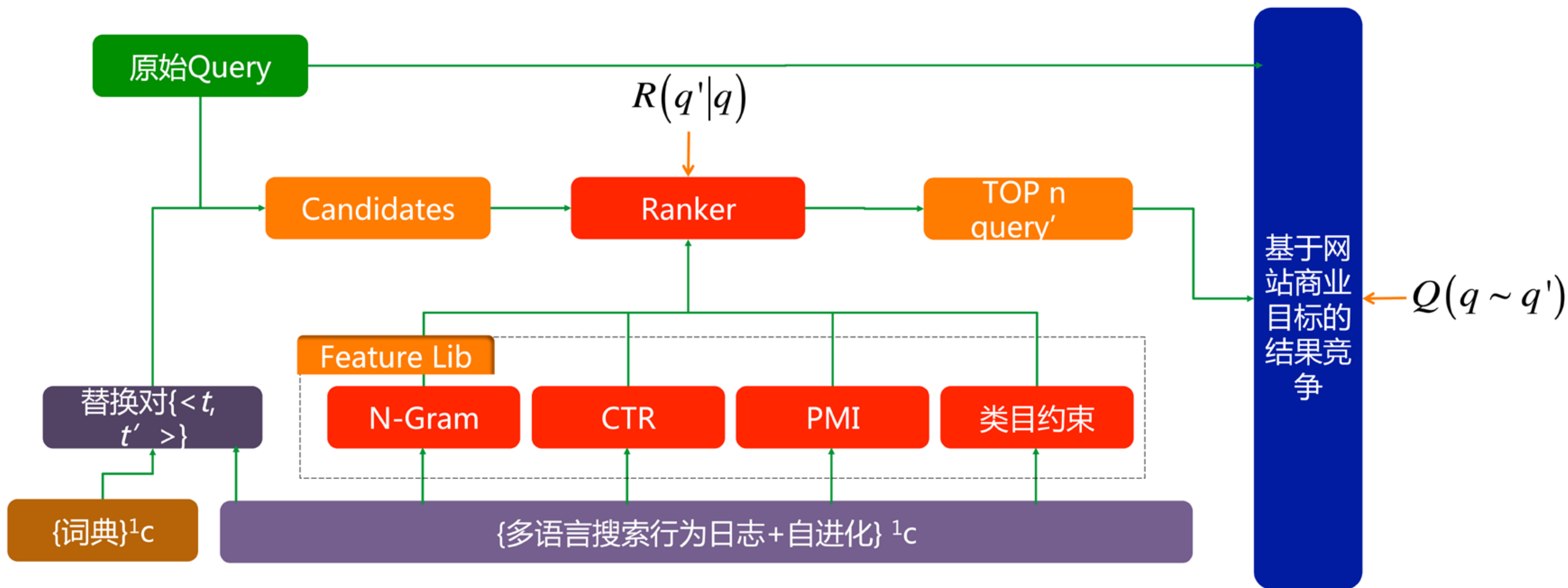
T_1

⋮

T_n



搜索&广告Query分析框架

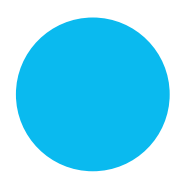


● 计算目标：

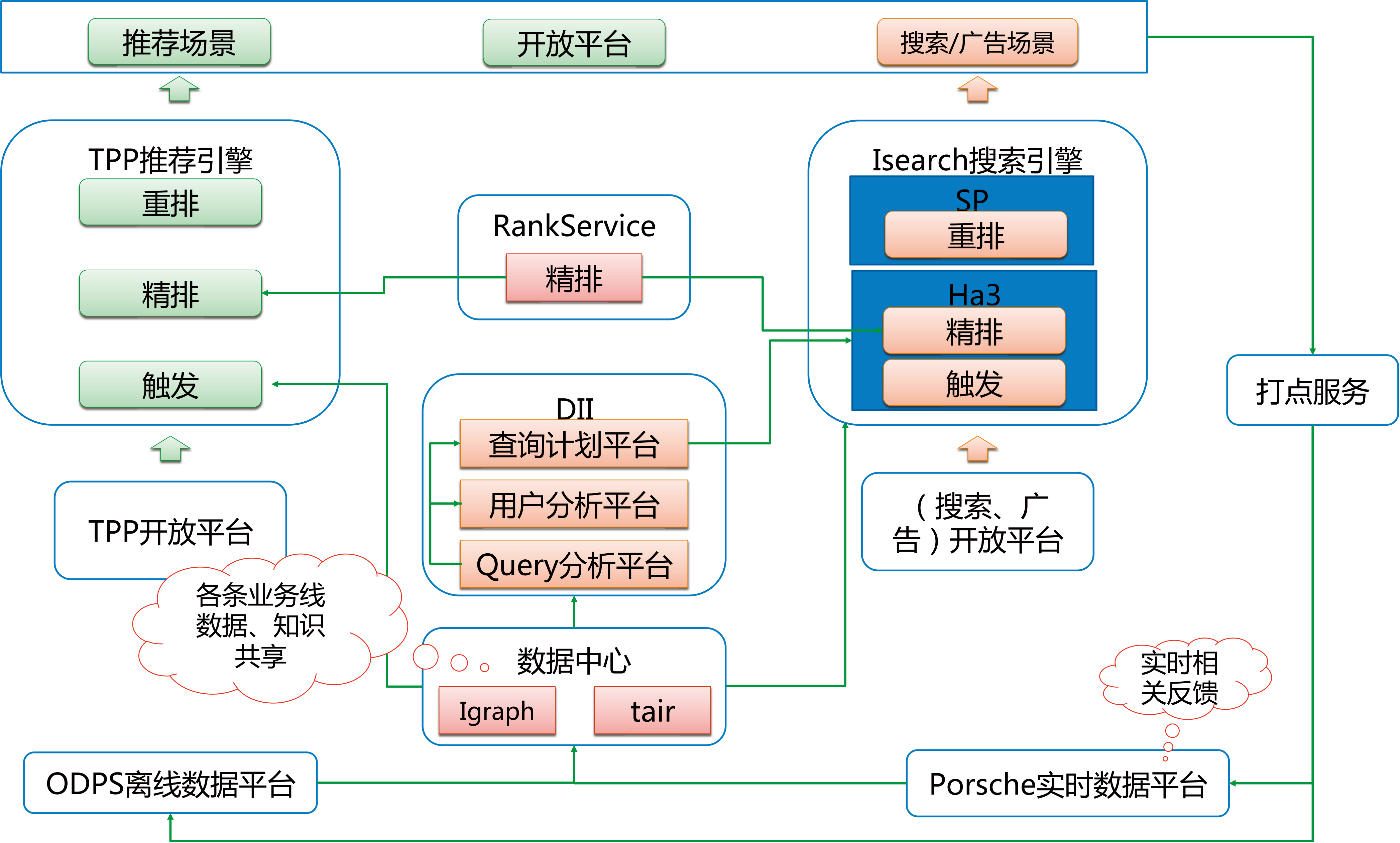
$$S(q \rightarrow q') \propto R(q'|q)Q(q' \sim q)$$

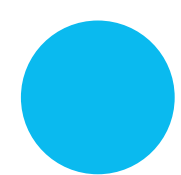
变换相关性

网站商业目标



搜索、广告、推荐一体化





跨境电商的风控特点

险

- 买家端、卖家端都有存在损害他人利益的风险，破坏平台秩序，累积有害数据。

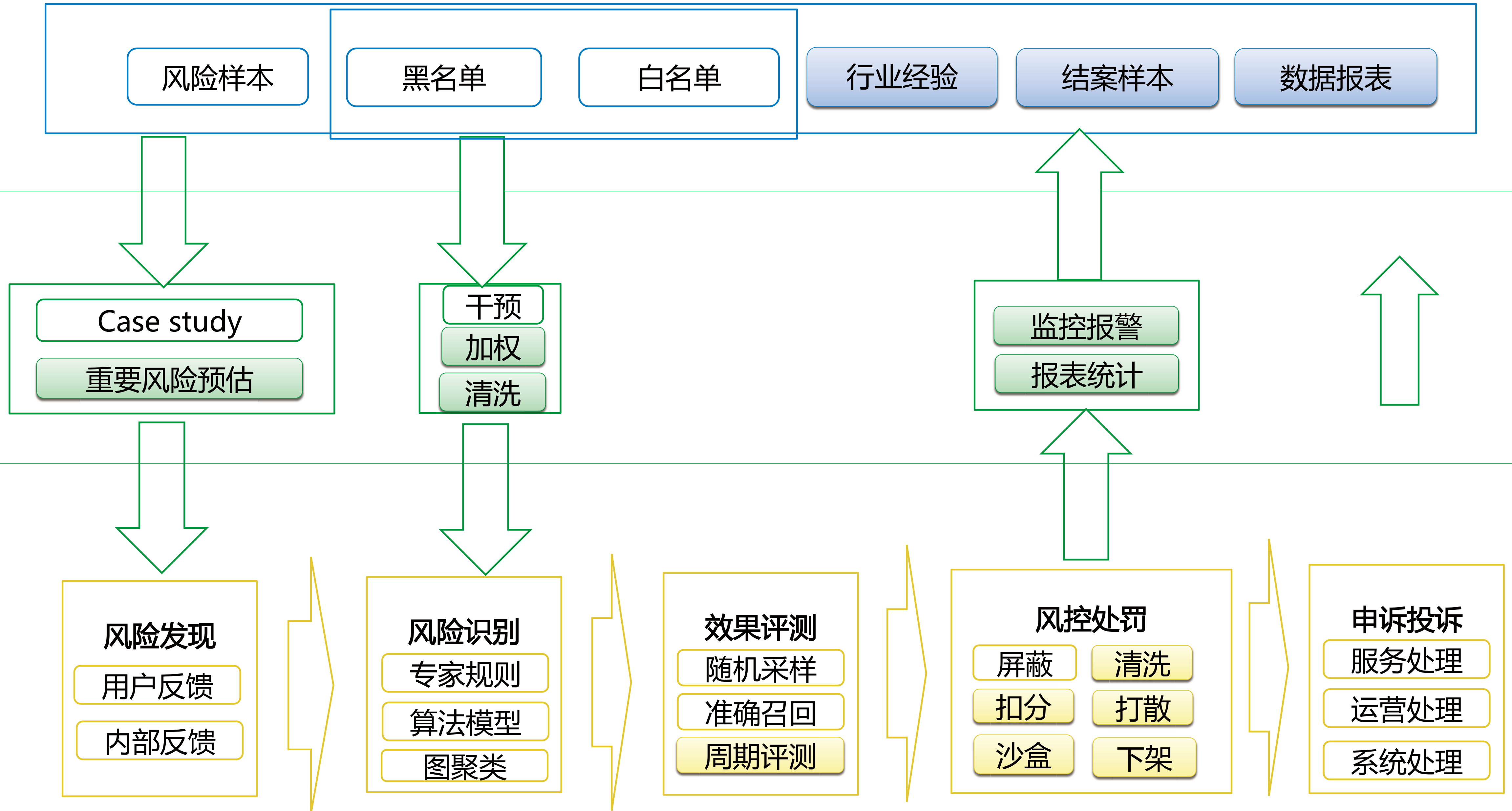
多

- 危害类型繁杂，危害手段和场景很多，往往道高一尺，魔高一丈，需要对业务有较强的理解。

急

- 紧急、突发情况非常多，有时差，需要更主动的防御。

风控机制建设(买家端&卖家端)



目录



● 多语言自然语言处理(NLP)技术

业务层

相关性

意图分析

推荐

导航

文本分类告

相关性

推词

意图分析

SEO

应用层

智能推荐

查询纠错

类目预测

语种识别

智能翻译

相关推荐

查询扩展

短标题优化

文本质量诊断

运营&评测

基础算法

长串

分句

依存句法

文本分类

文本聚类

主题模型

Text2Vec

短语

拼写检查

短语抽取

Term Weight

中心词

语言模型

Phrase2vec

词

Stemming

新词发现

分词/词性

专名识别

Word
Embedding

迁移学习

数据层

Domain-dependent

商品类目

同义词

依存分析语料

风格词

商品标题

主配件

双语品牌

上下位词

商品属性

产品词

品牌型号

材质词

Domain-independent

公司

口语书面语

国家词

短语库

常用词

化学词

比较词

单位词

界面展示

运营干预

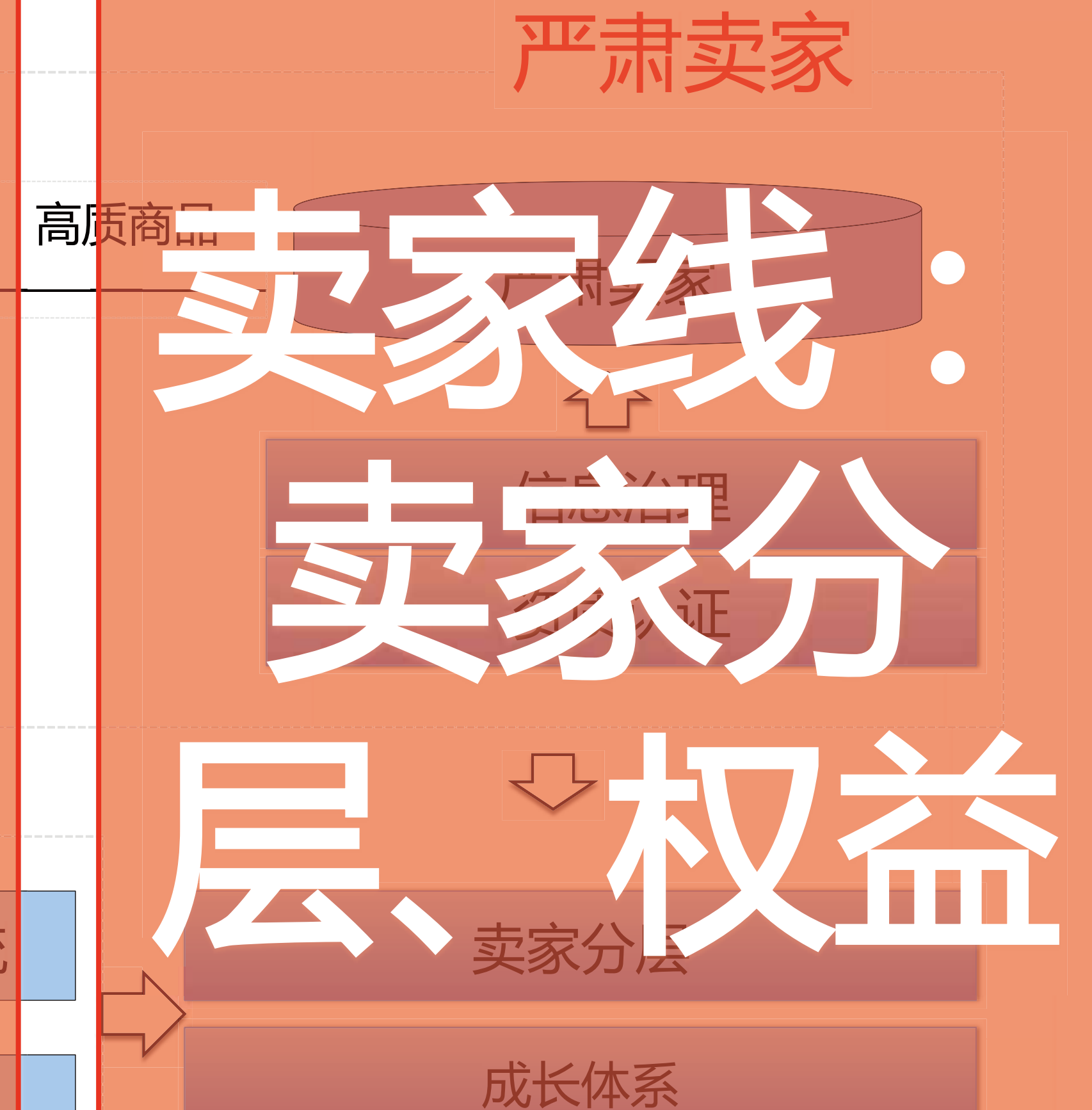
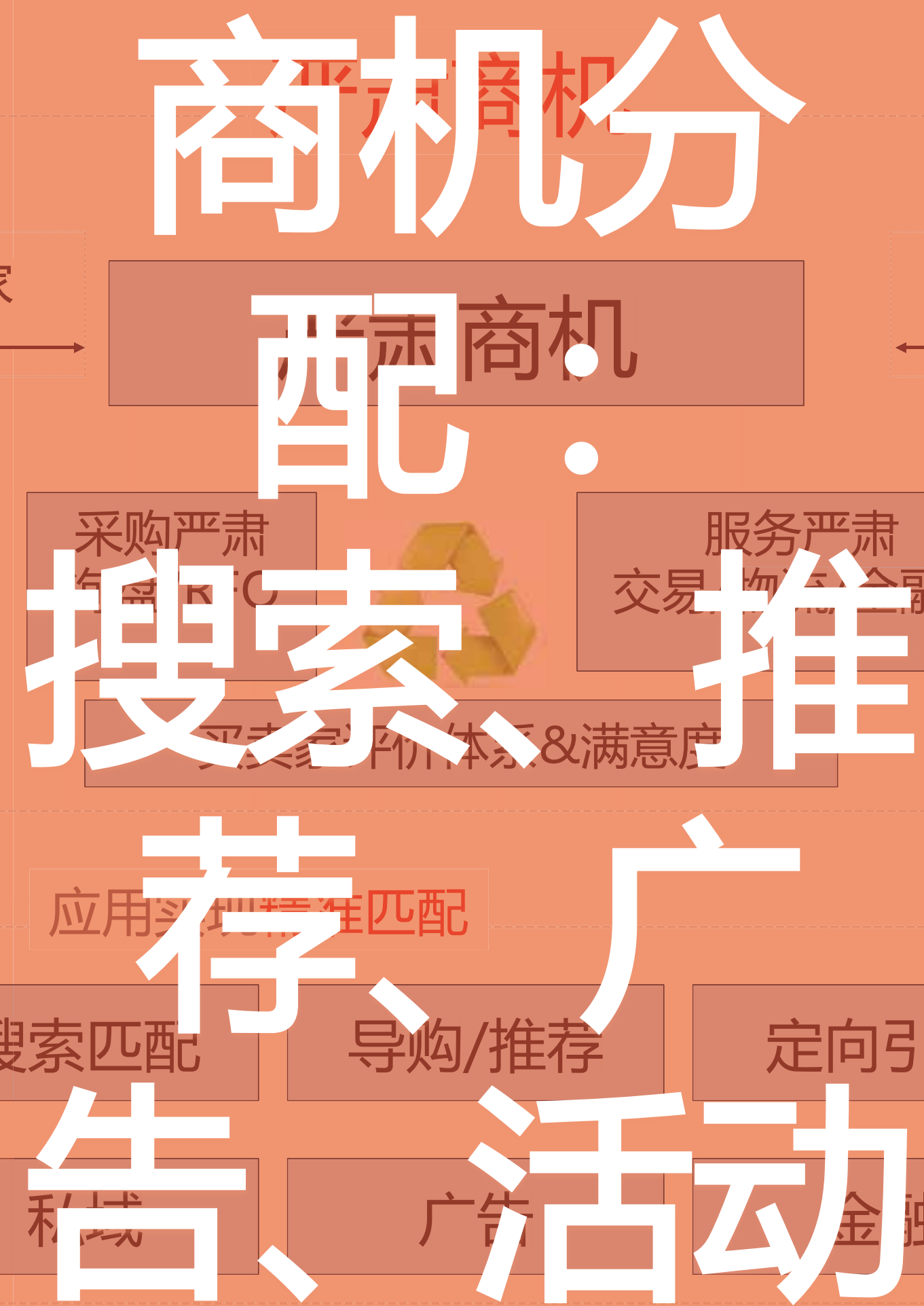
在线评测

评测报告

数据管理

目录





感谢你对阿里技术的关注。



微信扫描二维码
关注阿里技术



钉钉扫描二维码，或搜索群号
21764737，加入读者交流群