

# 阿里2B电商核心问题以及算法建模

---



【可发布】2019 QCon北京-洛俊宇-百度搜索引擎中的工程实践-百度-曹皓.pdf

霍承富/晓惑

阿里巴巴-CBU技术部-新零售算法

# 想做团队的领跑者 需要迈过这些“槛”

成长型企业，易忽视人才体系化培养  
企业转型加快，团队能力又跟不上

VS

从基础到进阶，超100+一线实战  
技术专家带你系统化学习成长

团队成员技能水平不一，  
难以一“敌”百人需求

VS

解决从小白到资深技术人所遇到  
80%的问题

寻求外部培训，奈何价更高且  
集中式学习

VS

多样、灵活的学习方式，包括  
音频、图文 和视频

学习效果难以统计，产生不良循环

VS

获取员工学习报告，查看学习  
进度，形成闭环

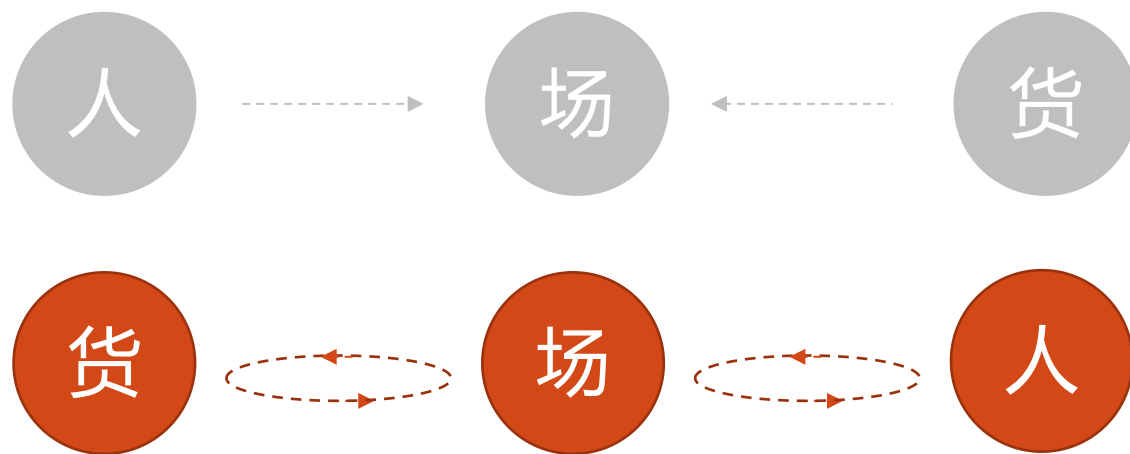


课程顾问「橘子」

回复「QCon」  
免费获取  
学习解决方案

# 极客时间企业账号 # 解决技术人成长路上的学习问题

# 电商核心问题



算法构建三位一体营销体系

## 问题定义：

- 如何挖掘人的诉求？如何表达场的心智？如何圈选符合心智的商品？
- 如何构建人货场的精准匹配机制？

# 算法建模

## 基础信息挖掘

- 人的表征
- 货的画像
- 场的图谱

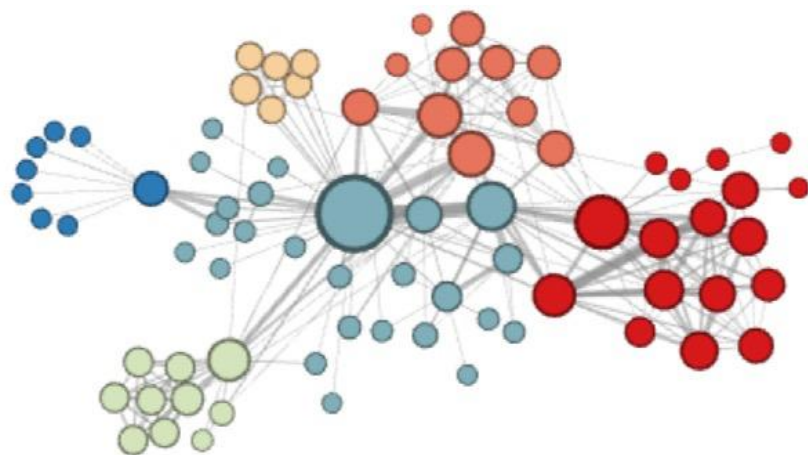
## 市场牵引机制

- 流量中控
- 召回排序
- 组货推荐

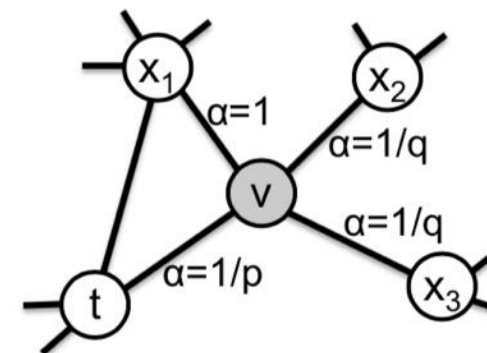
实时数据服务平台

# 算法建模 —— 人的表征

表征：非明文标签，基于深度模型的Embedding Vector。



- DeepWalk
- Node2Vector



$$\max_{\Phi} \sum_{v \in V} \log P(N(v) | \Phi(v))$$



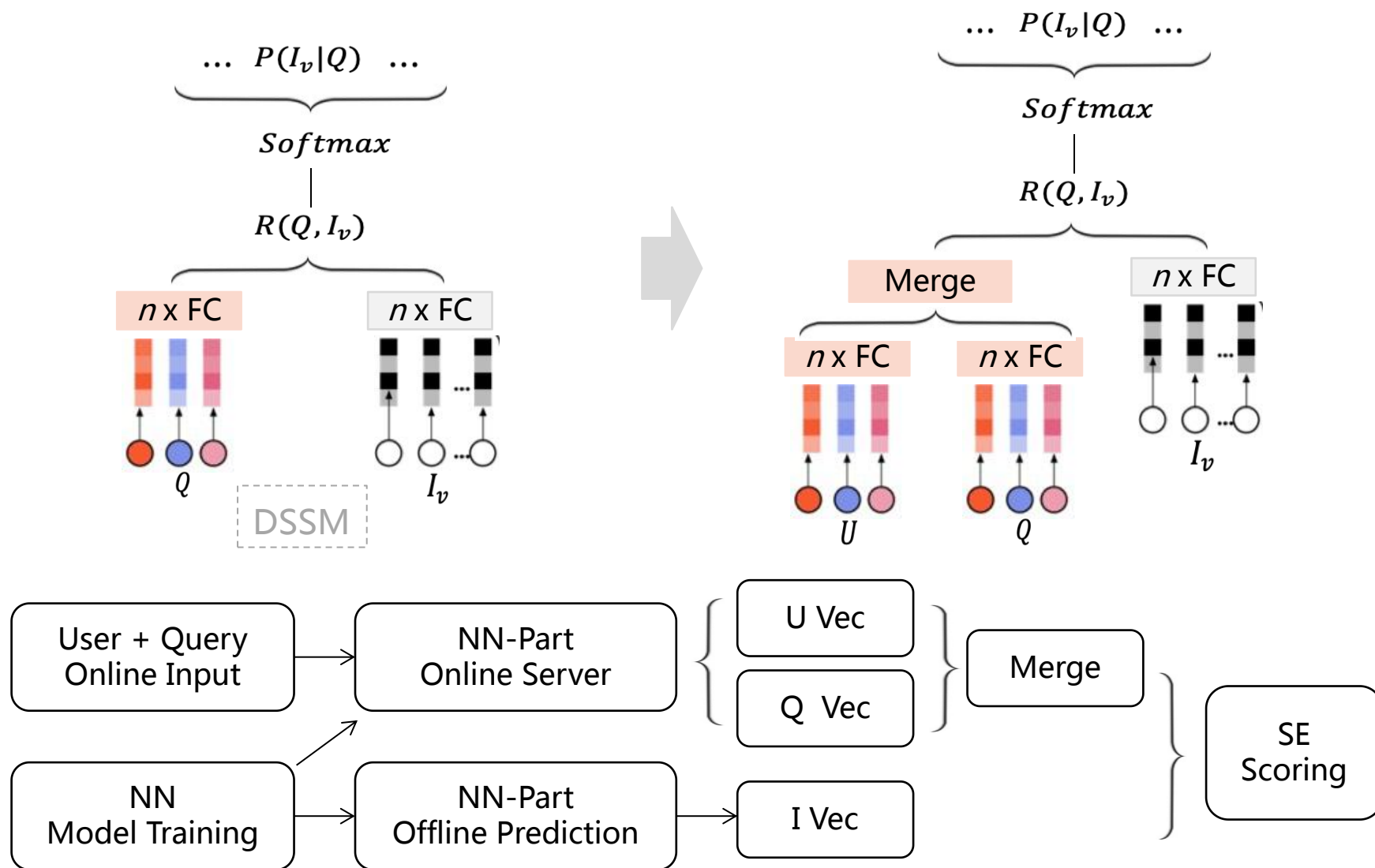
$$Vec_u = \sum_v Vec_{I_v}$$

• I2I • U2I • U2U

**问题：**基于全局行为Item网络的Embedding Vector缺乏精细的个性化刻画能力。

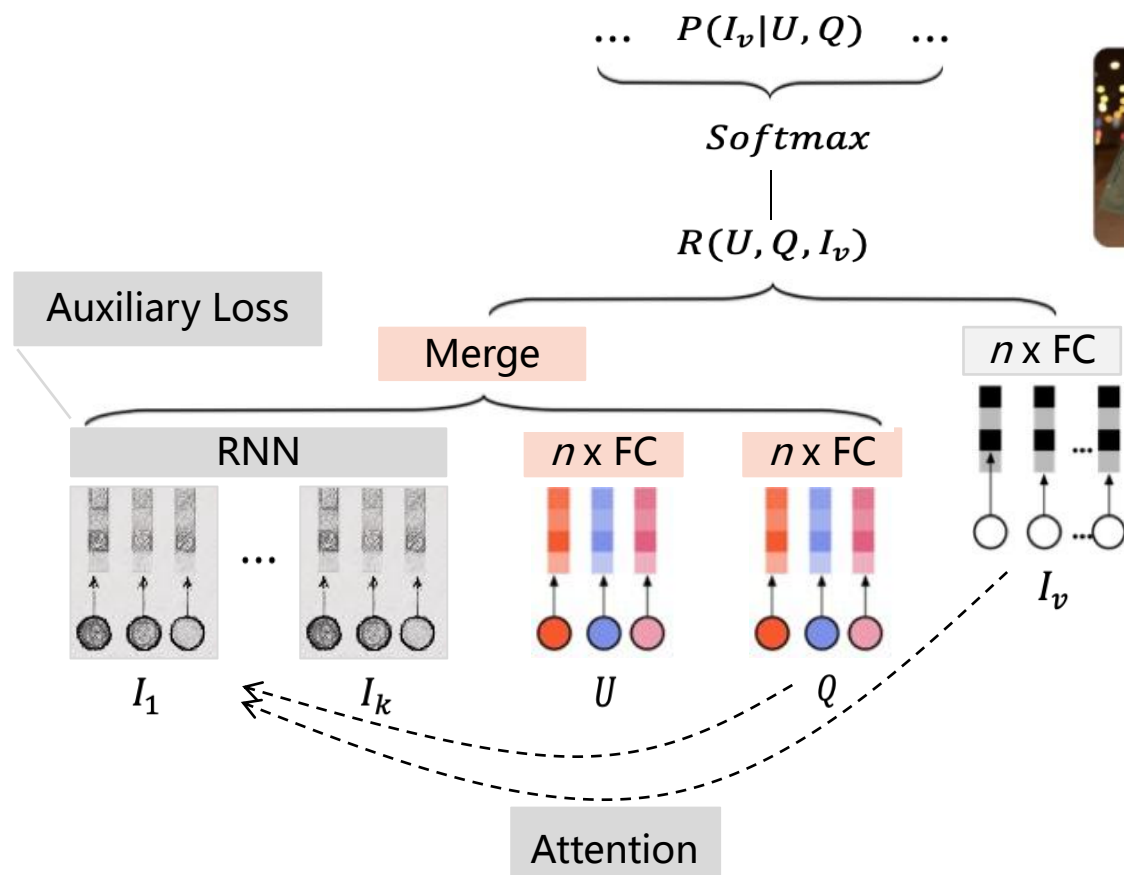
# 算法建模 —— 人的表征

**Opt1** : 上下文相关的Embedding Vector ( 个性化 )。



# 算法建模 —— 人的表征

Opt2 : 上下文相关+兴趣点演进的Embedding Vector。



- Feature Engineering
- Loss Function

$$L = L_o + L_a$$

- DIN ( Deep Interest Network )
- DIEN ( Deep Interest Evolution Network )



# 算法建模 —— 货的画像

画像：产品、运营、商家可感知的明文标签，可理解、可运营、可执行。



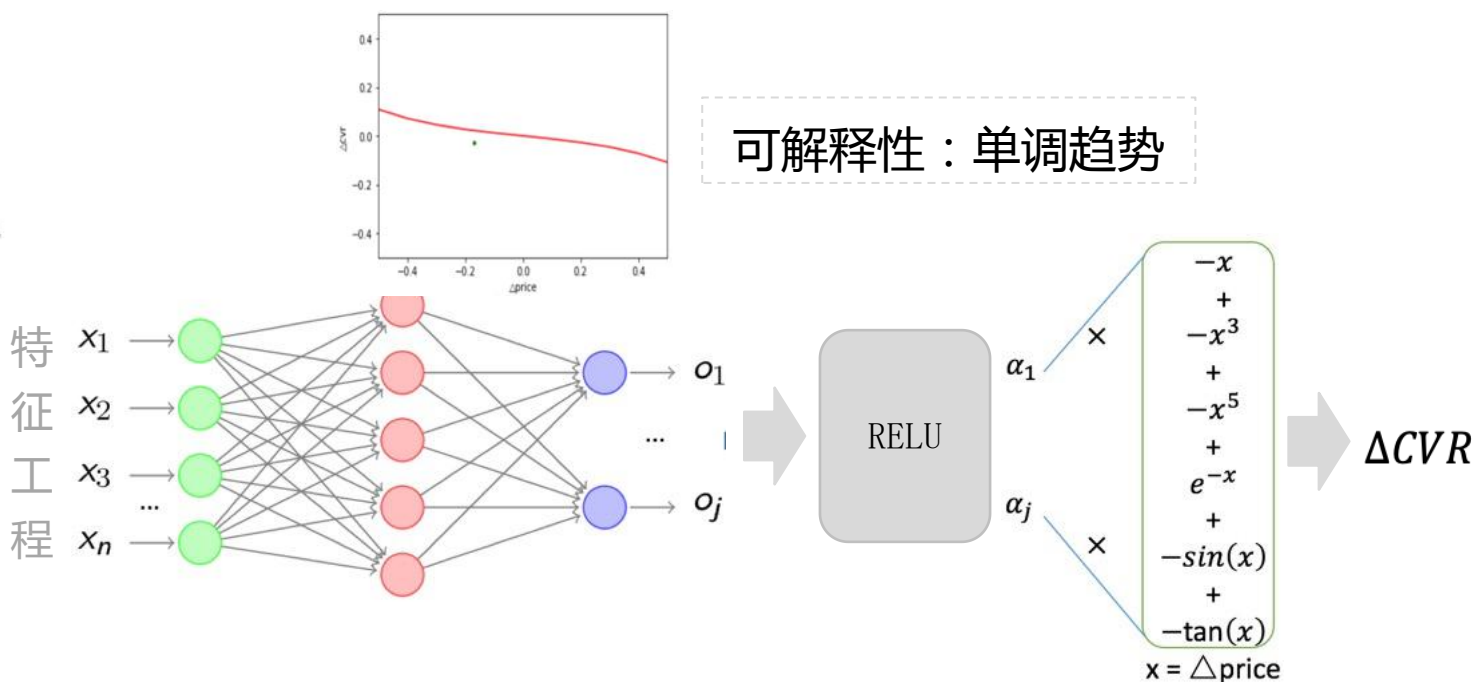
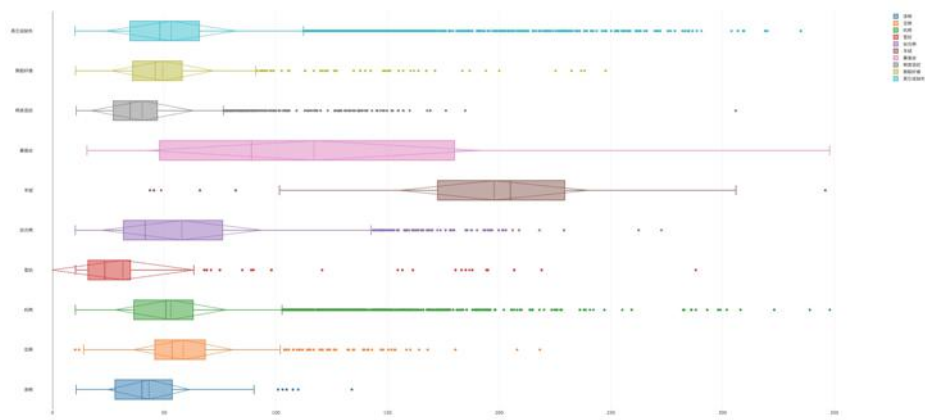
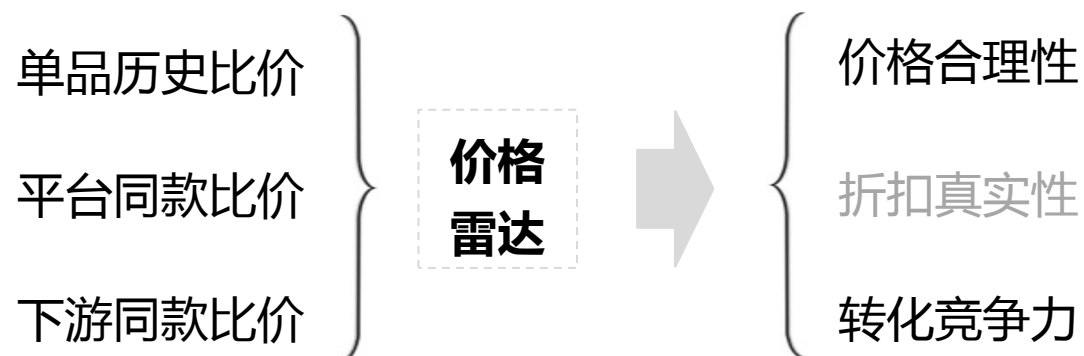
## 品类规划

- 价格机制：维护健康的价格区间，挖掘价格和转化的关系，建模价格竞争力。
- 销量预估：预估销量趋势，结合运营知识，定义生命周期。



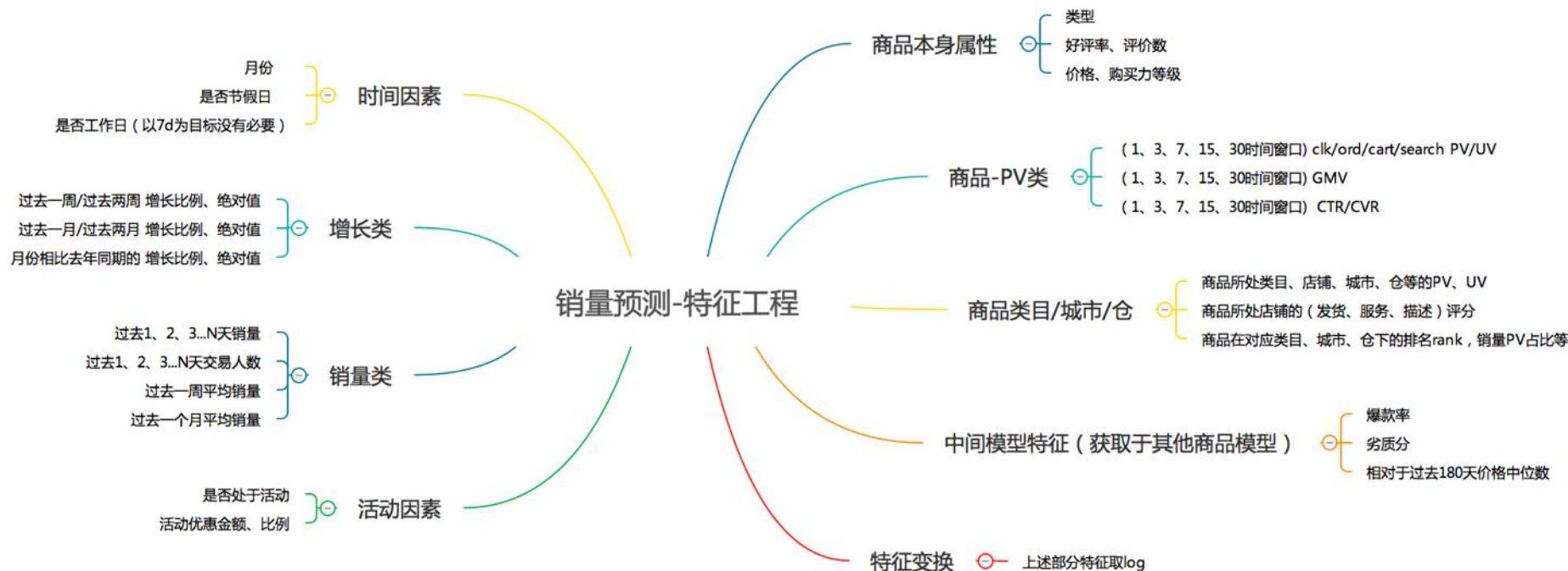
# 算法建模 —— 货的画像

## 价格机制



# 算法建模 —— 货的画像

## GBDT/DNN销量预估



| Item_id     | 过去7天销量 | 未来7天真实销量 | 未来7天预测销量 | case结论 |
|-------------|--------|----------|----------|--------|
| 565****4400 | 208    | 2596     | 2296     | 发现爆款   |
| 565****5247 | 37     | 3345     | 5060     | 发现爆款   |
| 546****5744 | 894    | 1234     | 1181     | 预测准确   |

- 历史规则圈选召回/准确 57%
- 算法预估圈选召回/准确 90%

# 算法建模 —— 场的图谱

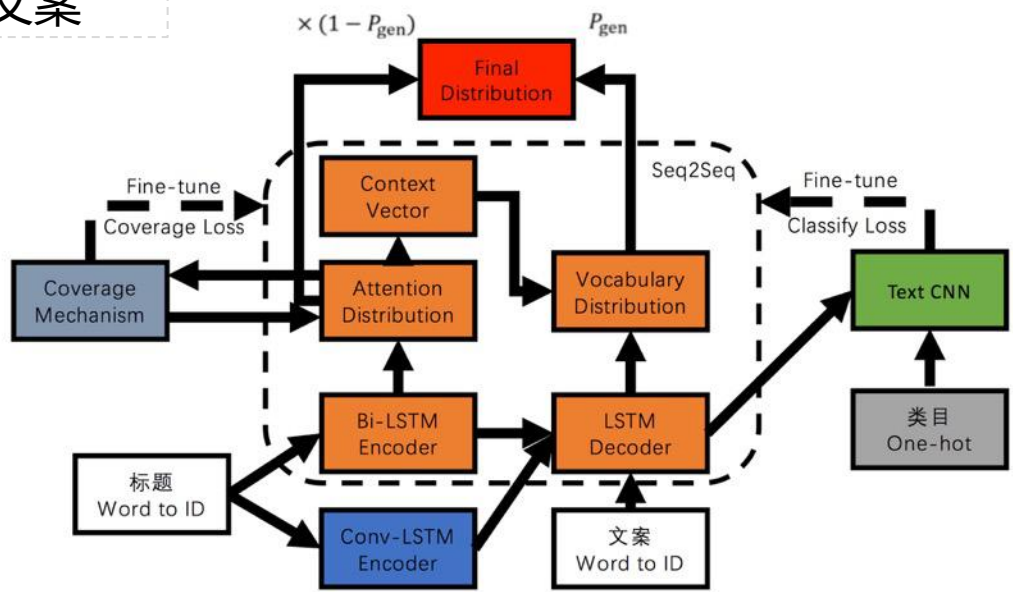
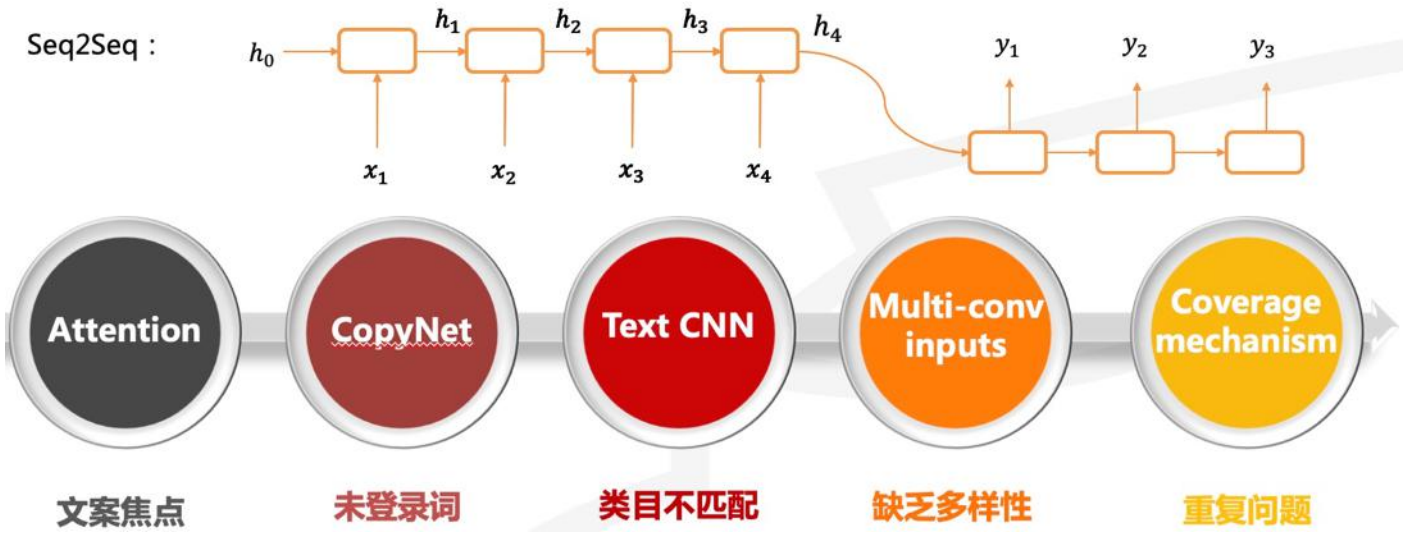
图谱：关系链接，具备心智表达能力和推理能力。



Clustering

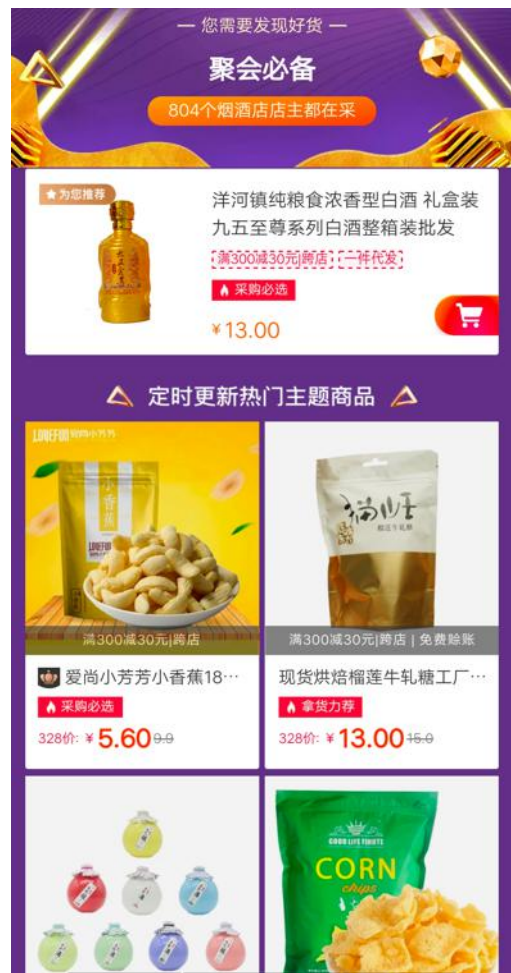
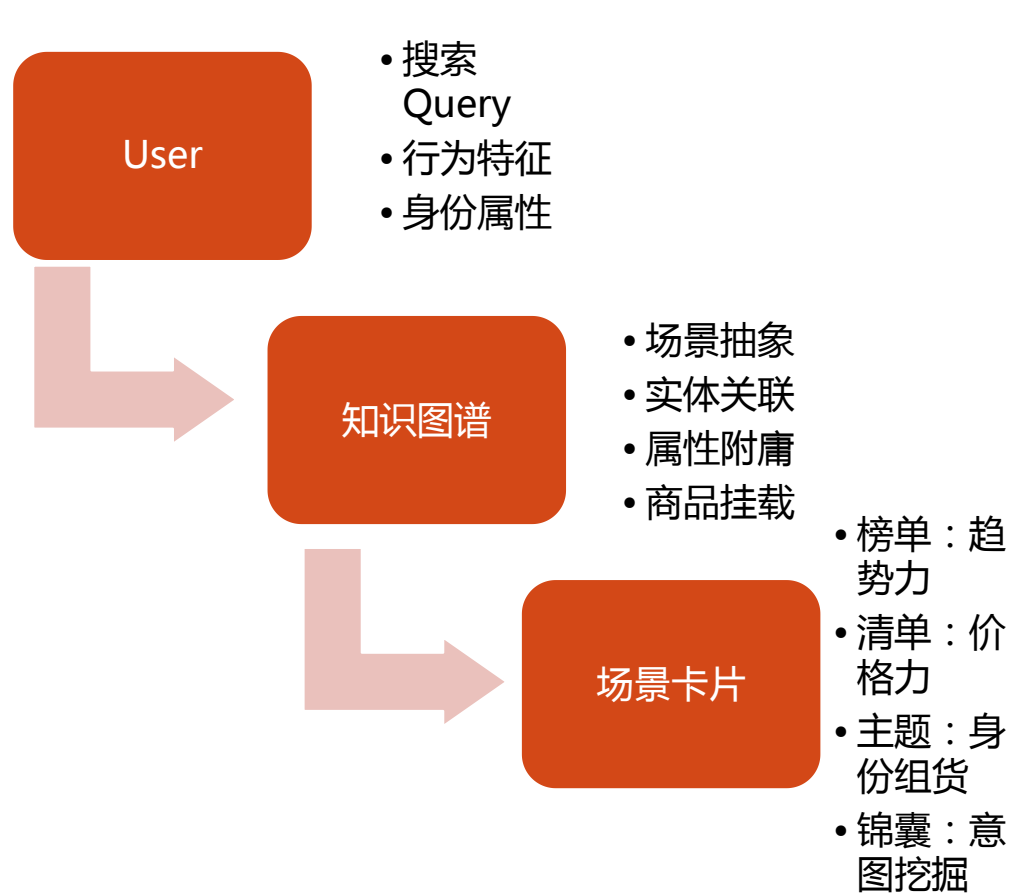


Scene2Scene



# 算法建模 —— 场的图谱

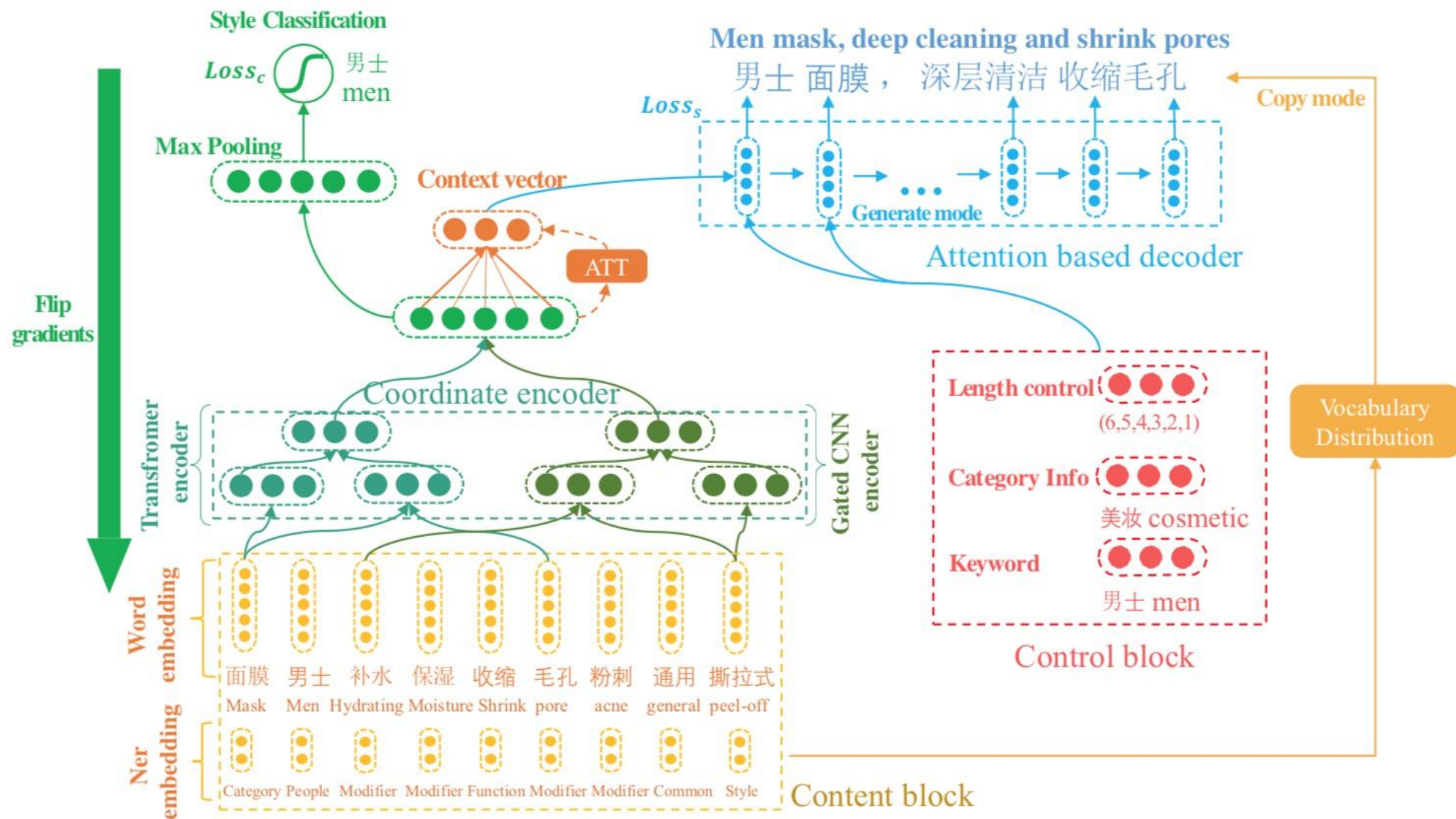
心智选品：基于场景心智，构建差异性的Clustering选品算法体系





# 算法建模 —— 场的图谱

营销文案：满足不同业务场景对心智风格和文案长度的强诉求。（WWW 2019）



# 算法建模 —— 场的图谱

## 营销文案：Case展示

### 风格控制

- 输入端添加风格词的embedding和NER信息
- 风格词和输入词语共享embedding信息

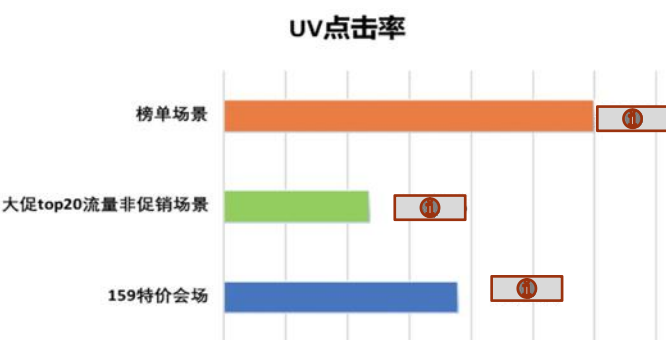
| 标题                        | 风格  | 文案                    |
|---------------------------|-----|-----------------------|
| 羊毛呢 毛领 时尚 斐女大衣 大码 外套 厚 连帽 | 爆款  | 这才是 爆款 的正确 打开 方式。     |
|                           | 折扣  | 年末 折扣 大作战， 毛领 大衣 来 帮你 |
|                           | 简约  | 连帽 棉服， 简约 不 单调 的 穿搭   |
|                           | 小公主 | 大毛领 外套， 做个 温暖 系 的 小公主 |

### 长度控制

- 在输入端添加位置的embedding信息

|           |  |
|-----------|--|
| 商品标题      | 阳台 户外 藤编家具 别墅花园 组合 休闲 圆床 藤条 躺床 沙发              |
| 词数：5（含标点） | 复古 的 中式 风格。                                    |
| 词数：10     | 这 款 休闲 圆床 藤条 藤编家具 组合 休闲 沙发。                    |
| 词数：18     | 经典 的 中式 传统 风格 的 一 款 沙发， 整体 造型 简约 大气， 非常 适合 小户型 |
| 词数：19     | 经典 的 中式 传统 风格 的 一 款 沙发， 整体 造型 设计 简约 大气， 非常 百搭。 |
| 词数：20     | 经典 的 中式 传统 风格 的 一 款 沙发， 整体 造型 简约 大气， 彰显 时尚 品味。 |

## 场景：榜单Case展示



## 基础信息挖掘

- 人的表征
- 货的画像
- 场的图谱

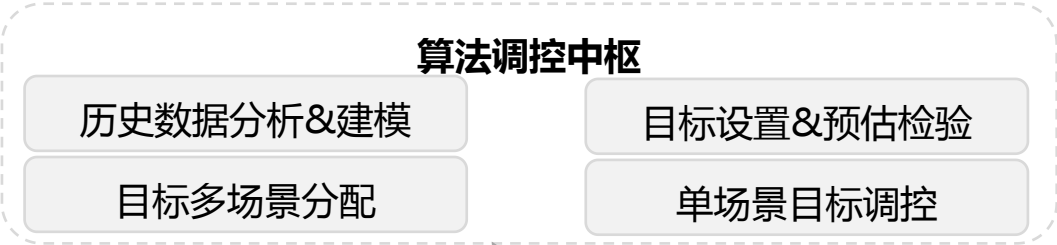
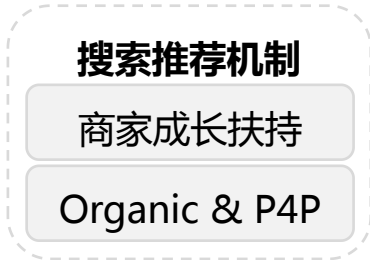
## 市场牵引机制

- 流量中控
- 召回排序
- 组货推荐

实时数据服务平台



# 算法建模 —— 流量中控



# 算法建模 —— 流量中控

**Organic & P4P**：平衡广告营收和买家体验，最大化平台价值。

$$RankScore_{p4p} = pCTR^t \cdot bidprice$$

$$RankScore_{org} = \alpha \cdot pCTR + \beta \cdot pCVR + \gamma \cdot REL + \dots$$

$$\max \sum_{i,j} REV_{ij} \cdot x_{ij} \quad \text{s.t.} \quad \frac{\sum_{i,j} CY_{ij} \cdot x_{ij}}{\sum_{i,j} x_{ij} \cdot k_{ij}} \geq T, \quad \sum_j x_{ij} = 1, \forall i, \quad x_{ij} \in \{0,1\}, \forall i,j.$$

$$\min \sum_i \beta_i \quad \text{s.t.} \quad \beta_i \geq REV_{ij} + \alpha \cdot (CY_{ij} - T \cdot k_{ij}), \quad \forall i,j, \quad \alpha \geq 0$$

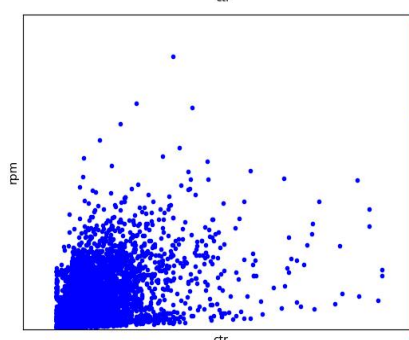
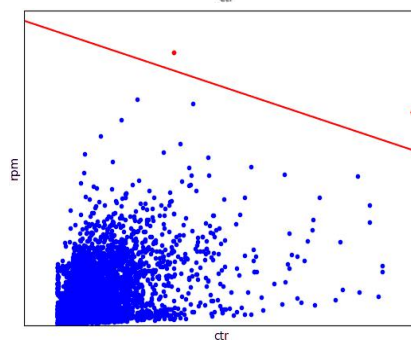
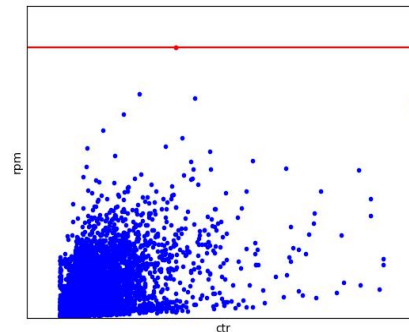
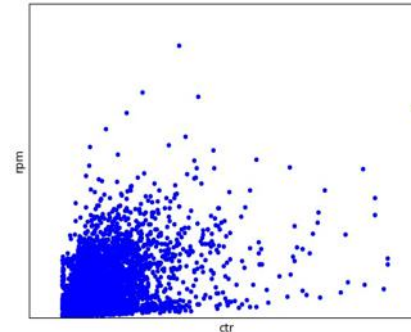
$$j^* = \operatorname{argmax}_j \{REV_{ij} + \alpha \cdot (CY_{ij} - T \cdot k_{ij})\}$$

$$\max \sum_{m,n} y_{mn} \cdot rpm_m \cdot pos_n + \alpha \cdot \left( \sum_{m,n} y_{mn} \cdot ctr_m \cdot pos_n - T \cdot \sum_{m,n} y_{mn} \right)$$

$$\text{s.t.} \quad \sum_m y_{mn} \leq 1, \quad \forall n,$$

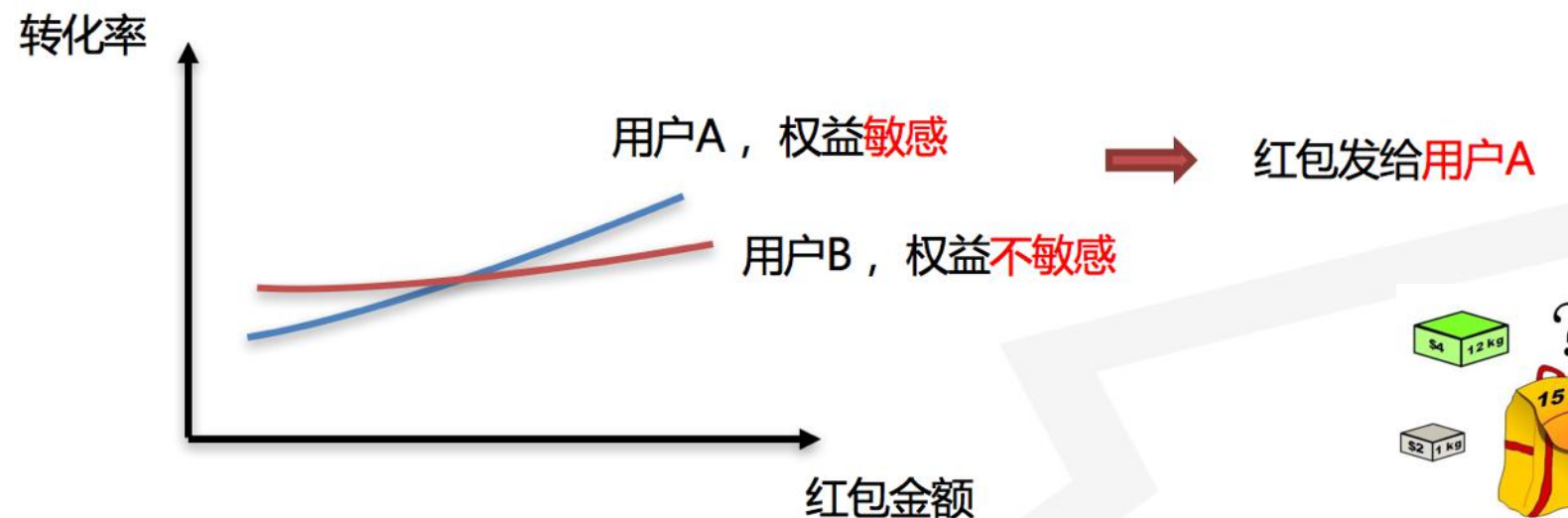
$$\sum_n y_{mn} \leq 1, \quad \forall m,$$

$$\sum_{m,n} y_{mn} \leq K, \\ y_{mn} \in \{0,1\}, \quad \forall m,n.$$



# 算法建模 —— 流量中控

**权益分发：**用户权益敏感度预估+权益金额决策分配（不同的用户分发不同额度的权益，最大化业务目标）



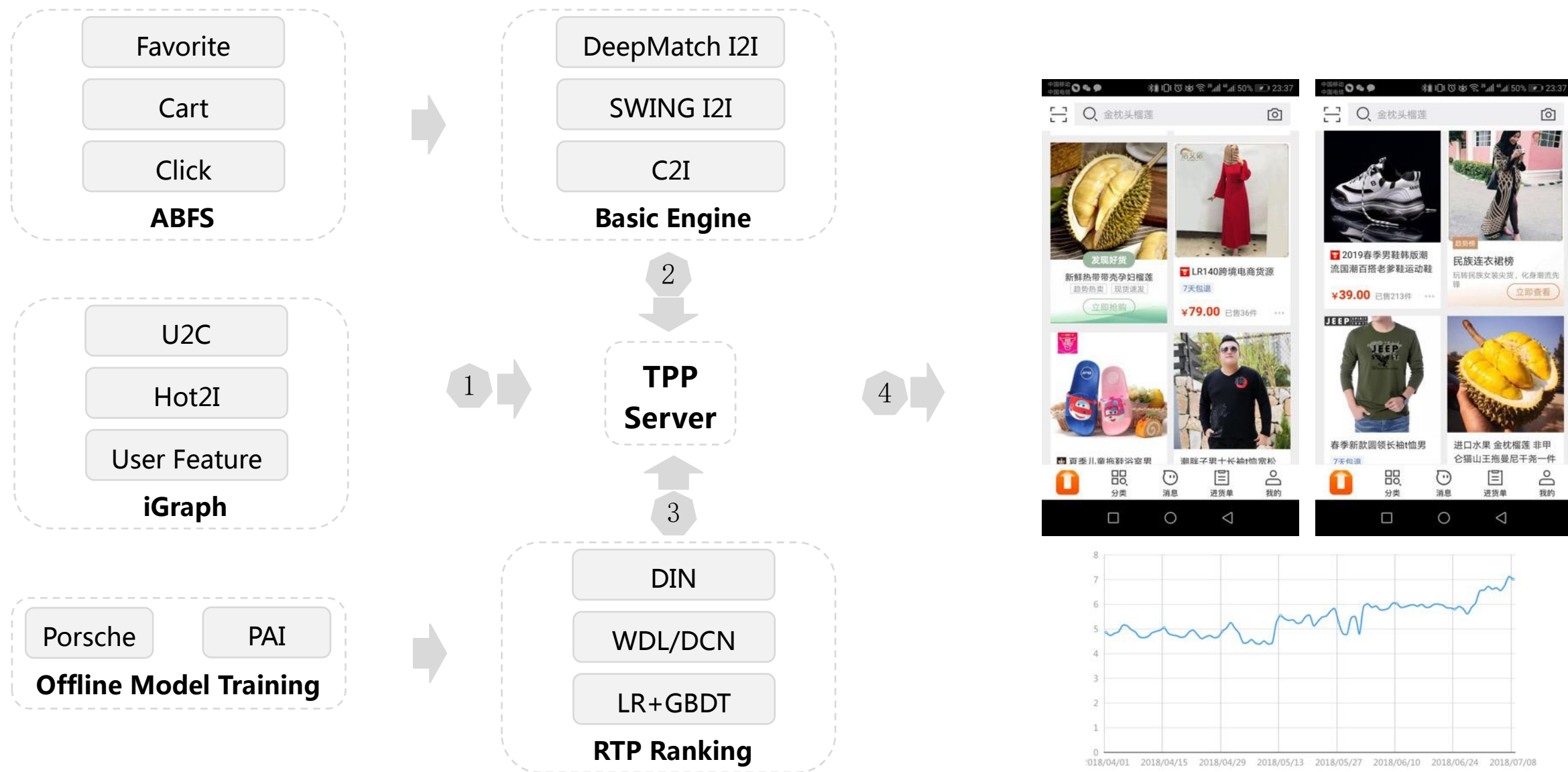
## 分组背包MCKP

- M用户（M组）
- N种额度权益（N物品）
- P权益敏感概率
- W权益额度/权益预算

$$\begin{aligned} \max & \sum_{i=1}^M \sum_{j=1}^{N_i} p_{ij} x_{ij} \\ \text{s.t.} & \sum_{i=1}^M \sum_{j=1}^{N_i} w_{ij} x_{ij} \leq W \\ & \sum_{j=1}^{N_i} x_{ij} = 1, i = 1, \dots, M \\ & x_{ij} \in \{0, 1\}, i = 1, \dots, M, j = 1, \dots, N_i \end{aligned}$$



# 算法建模 —— 召回排序



# 算法建模 —— 召回排序

## 大促会场实时个性化：业务强诉求

- 会场商品实时圈选投放和下架
- 售罄商品实时下架
- 实时指标（销量/转化率/综合）排序



- User \* Item Hit统计特征
- Item统计特征
- 算法Join特征回传埋点

# 算法建模 —— 组货推荐

## 业务场景

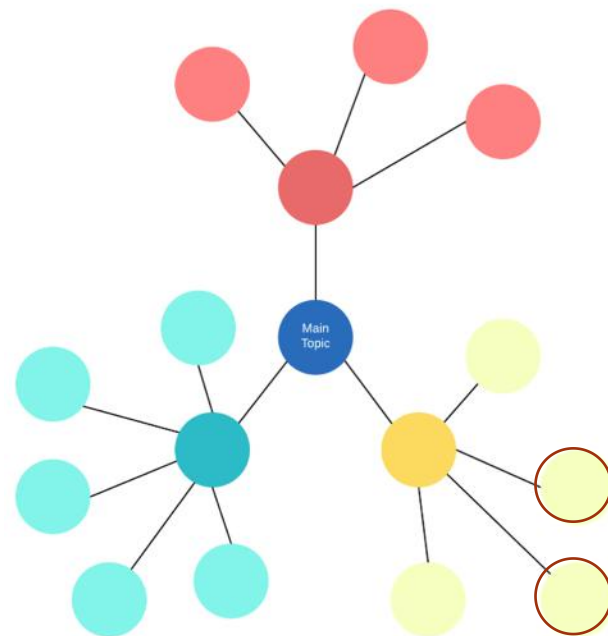
- 小B采购 ( e.g. 渔具商行 )
- 类C自购 ( e.g. 钓鱼套装 )
- 营销活动 ( e.g. 满减、搭配推荐促GMV )

## 运营方案

- 主营类目圈品
- 基于类目树组货
- 行业知识穷举

## 算法方案

- 用户行为网络图聚类
- 金额约束规划
- 行业知识约束规划





# 算法建模 —— 组货推荐

LPA : each node is initialized with unique label and at every iteration of the algorithm, each node adopts a label that a maximum number of its neighbors have.

Initialization:  $C_x(0) = x$

Loop:

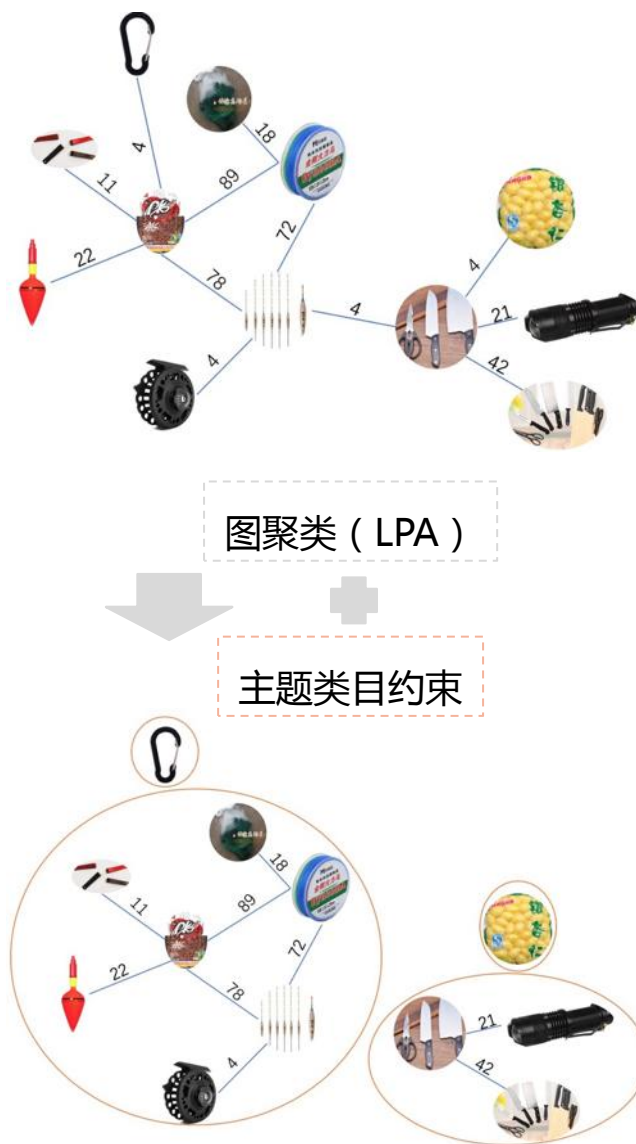
$$C_x(t) = f(C_{x,1}(t), \dots, C_{x,m}(t),$$

$$C_{x,m+1}(t-1), \dots)$$

if  $d_x^C \geq d_x^{C^*} \forall *$  and  $x$

break

**Note** : 抽象业务诉求，采用成熟的算法，并结合业务知识，求解业务问题。



**组货连拍**

推荐组合一

| 商品名称       | 价格        |
|------------|-----------|
| 六子渔具新品玻... | ¥19.90 ×1 |
| 鲢鳙球形电子漂... | ¥3.00 ×1  |
| 垂钓小配件、线... | ¥2.50 ×1  |

4件打包价: ¥14.00 (省¥14.60) **一键连拍**

今日最热 母婴玩具 日用百货 食品生鲜 美妆个护 家电数码 家纺

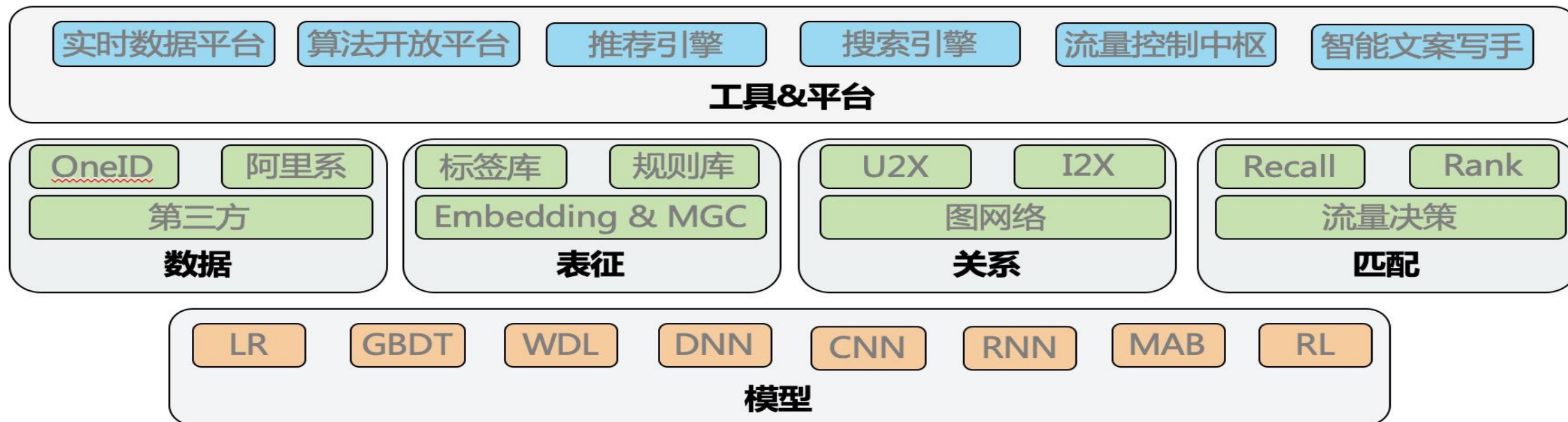
超值连拍5折起

| 商品名称                | 连拍价    | 限购   |
|---------------------|--------|------|
| 水晶绒披风秋冬款加厚连帽...     | ¥17.00 | 限购1件 |
| 澳乐不倒翁玩具0-3-6-12个... | ¥9.90  | 限购1件 |

**加入连拍单**



# 算法能力体系



- Tree-based Subgraph Convolutional Neural Networks for Top-N Recommender Systems, submitted, IJCAI 2019.
- Automatic Generation of Pattern-controlled Product Description in E-commerce, accepted, WWW 2019.
- User behavior sequence modeling to optimize ranking mechanism for E-commerce search, ICCIP 2017: 164-169.

# 极客邦科技 会议推荐2019

5月

**QCon** 北京

全球软件开发大会

大会: 5月6-8日  
培训: 5月9-10日

**QCon** 广州

全球软件开发大会

培训: 5月25-26日  
大会: 5月27-28日

6月

**GTLC**  
GLOBAL  
TECH LEADERSHIP  
CONFERENCE

上海

技术领导力峰会

时间: 6月14-15日

**GMTC** 北京

全球大前端技术大会

大会: 6月20-21日  
培训: 6月22-23日

7月

**ArchSummit** 深圳

全球架构师峰会

大会: 7月12-13日  
培训: 7月14-15日

10月

**QCon** 上海

全球软件开发大会

大会: 10月17-19日  
培训: 10月20-21日

11月

**GMTC** 深圳

全球大前端技术大会

大会: 11月8-9日  
培训: 11月10-11日

**AiCon** 北京

全球人工智能与机器学习大会

大会: 11月21-22日  
培训: 11月23-24日

12月

**ArchSummit** 北京

全球架构师峰会

大会: 12月6-7日  
培训: 12月8-9日

# 阿里2B电商核心问题以及算法建模

---

谢谢

Q & A