

爱奇艺效果广告的个性化探索与实践

原创 王晖 DataFunTalk 2019-10-17



爱奇艺效果广告的个性化探索与实践

王晖
爱奇艺 资深工程师



文章作者：王晖 爱奇艺 资深工程师

编辑整理：王浩

内容来源：爱奇艺技术沙龙

出品社区：DataFun

注：欢迎转载，转载请在留言区内留言。

导读：随着大数据和人工智能的技术进步，数字营销行业也在不断变化。得益于雄厚技术实力，爱奇艺效果广告得到蓬勃发展，实现了海量触达、定向精准、效果显著的广告投放，为广告主提供跨平台、多终端的网络推广服务。本次我将分享爱奇艺效果广告个性化方面的思考与实践，聚焦广告排序算法，希望能为大家带来一些启发。

包括以下三部分：

- 爱奇艺效果广告算法 - 背景&架构
- 点击&转化率预估 - 核心问题

- 智能出价 - 博弈&共赢

■ 爱奇艺效果广告算法-背景&架构

1. 广告资源位



如上图所示，爱奇艺主要的资源位可以用双引擎来概括，即信息流和框内：

- **信息流：**信息流广告位由准信息流和纯信息流共同构成的。其中，打开 App 显示的推荐页，由于是非原生信息，但是以信息流形式推送，所以我们称之为准信息流广告位。同时，爱奇艺平台提供了很多原生频道，比如热点频道，在这些频道下插入的广告称为纯信息流。
- **框内：**对于框内资源，效果广告主要投放在框内中插的位置上，同时也会做一些前贴和后贴的框内广告投放。
- **其他：**除上述两大引擎，还有猜你喜欢和视频关联位等资源位。

2. 计费方式

oCPX (optimized CPX) 业务是 oCPC (optimized Cost Per Click) 和 oCPM (optimized Cost Per Mille) 两种计费方式的统称，是从 CPX 业务演进而来，这里的 CPX 代表的是以点击计费和真实观看计费。

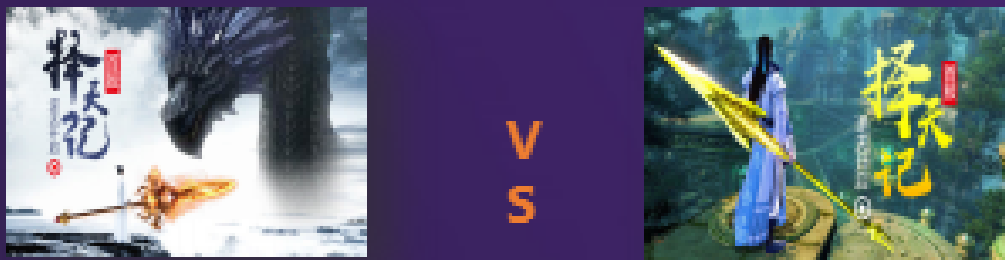


CPX 时代广告主投放广告是以点击作为计费方式，但广告主的真实诉求是点击后的转化行为。例如上图，一个 App 的广告，用户点击后的下载、安装、付费，这些转化行为才是广告主真正关注的。广告主为了提升从点击到转化的概率，会做很多离线优化，例如定向20-30岁女性或定向喜欢《亲爱的热爱的》电视剧的用户。但这样做可能会产生很多问题，比如定向粒度过细，目标人群少，导致广告放量效果不佳。

为解决上述问题，oCPX 的业务诞生了，优化目标从点击变成了后续的转化行为，广告主只需在投放广告时设置转化目标以及这个目标下的出价，其他的事情由算法模型来解决。

3. 算法难点

核心目标：优化转化效果



oCPX 业务对算法挑战难题主要为上图三点：

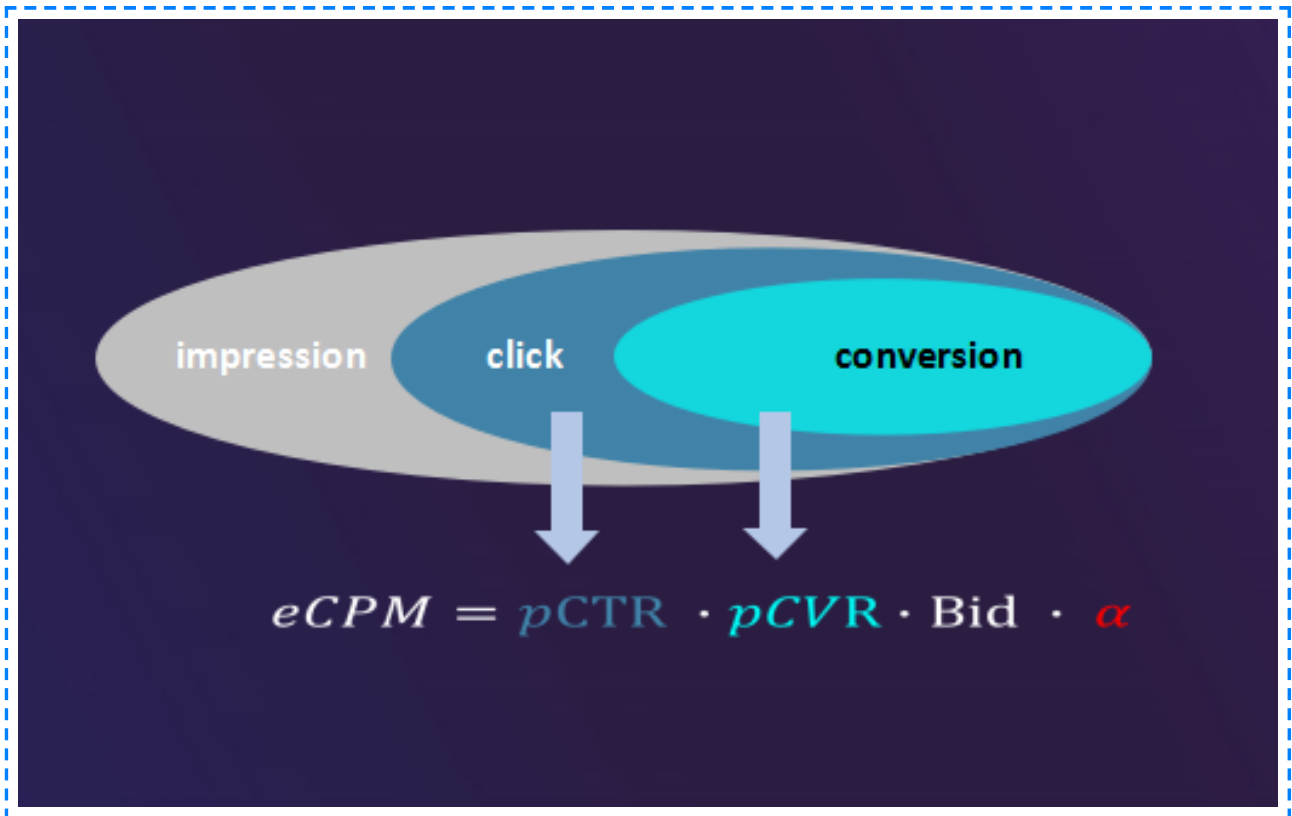
- 转化样本非常高维稀疏，样本维度高且正样本量极少。
- oCPX 业务复杂，比如需要支持安装转化、付费转化、关注公众号转化等类型。
- 平台流量大，对计算规模时效要求高。

4. 个性化广告推荐流程



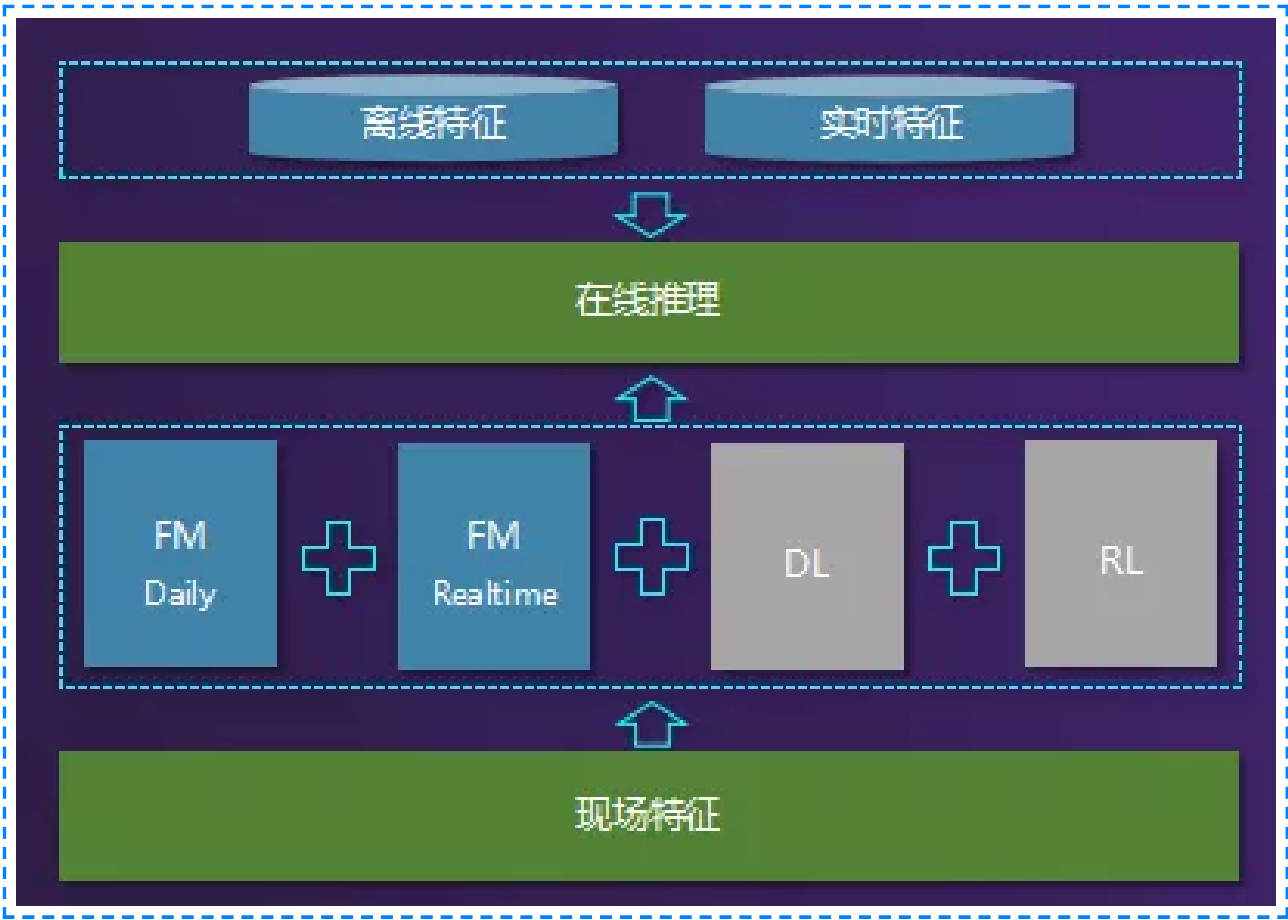
- 召回：对候选广告进行召回，比如人群定向等。
- 粗排：使用轻量级但保证精度的模型对广告做初选，避免精排阶段服务压力过大。粗排阶段还有冷启动和随机探索等功能模块，主要用于解决 E&E 问题，避免马太效应（投放好广告越来越好，差的越来越差）
- 精排：使用高精度模型对点击率、转化率和智能出价进行预估。另外精排阶段还支持预算平滑功能。还有预算平摊到全时段等。

5. 排序逻辑



粗、精排的排序逻辑依照 eCPM (effective Cost Per Mille)，也就是千次展示平台的预期收益来排序，其计算公式如上图所示， $eCPM = \text{点击率} \times \text{转化率} \times \text{出价} \times \text{智能出价因子}$ 。

6. 线上训练推理流程

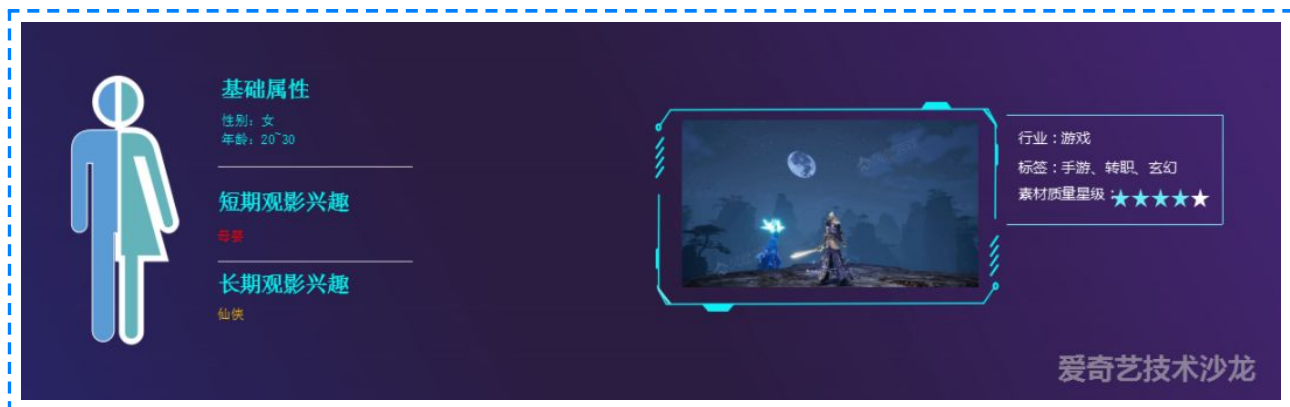


如上图自下而上：

- 1. 现场特征，相较离线特征，现场特征能真实还原线上环境。
- 2. 第二步进入模型训练，会有 FM 分天模型、FM 在线学习、深度学习和强化学习等。
- 3. 完成模型离线训练后推送至线上，结合离线、实时特征进行在线推理。

■ 点击&转化率预估-核心问题

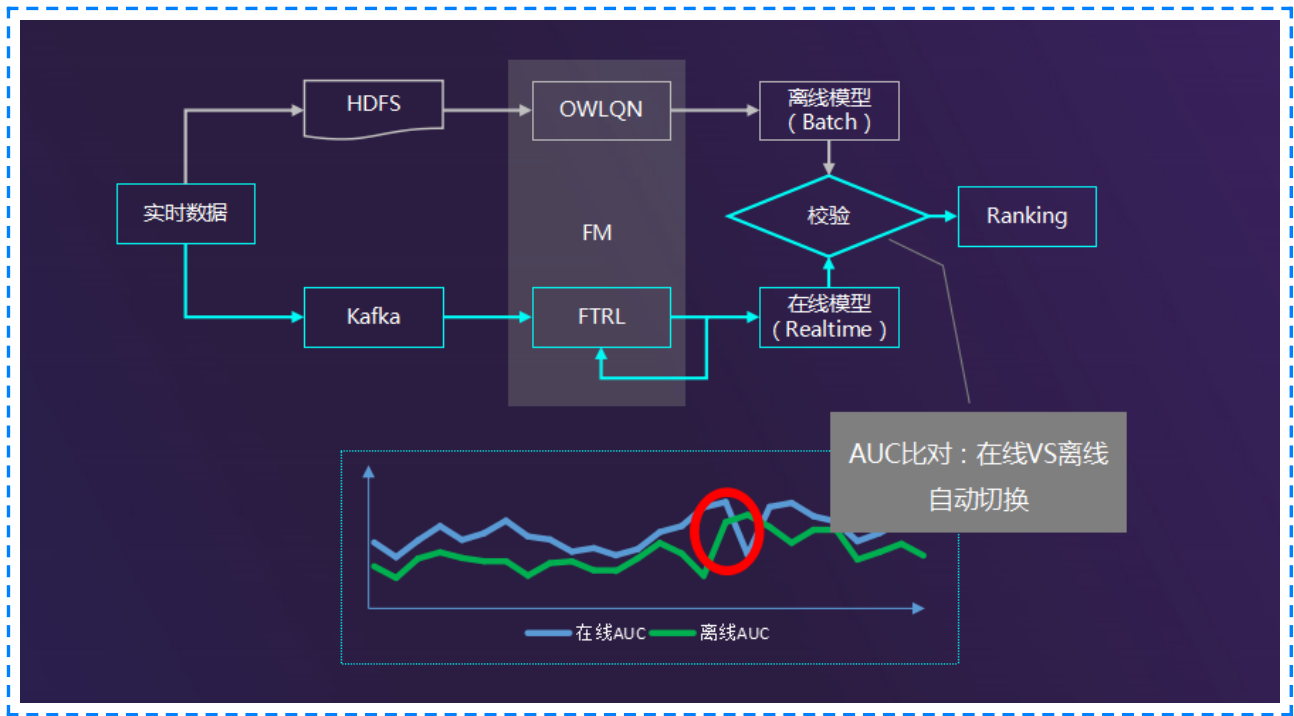
1. 特征工程



从周期来划分特征可分为实时、短期和长期。

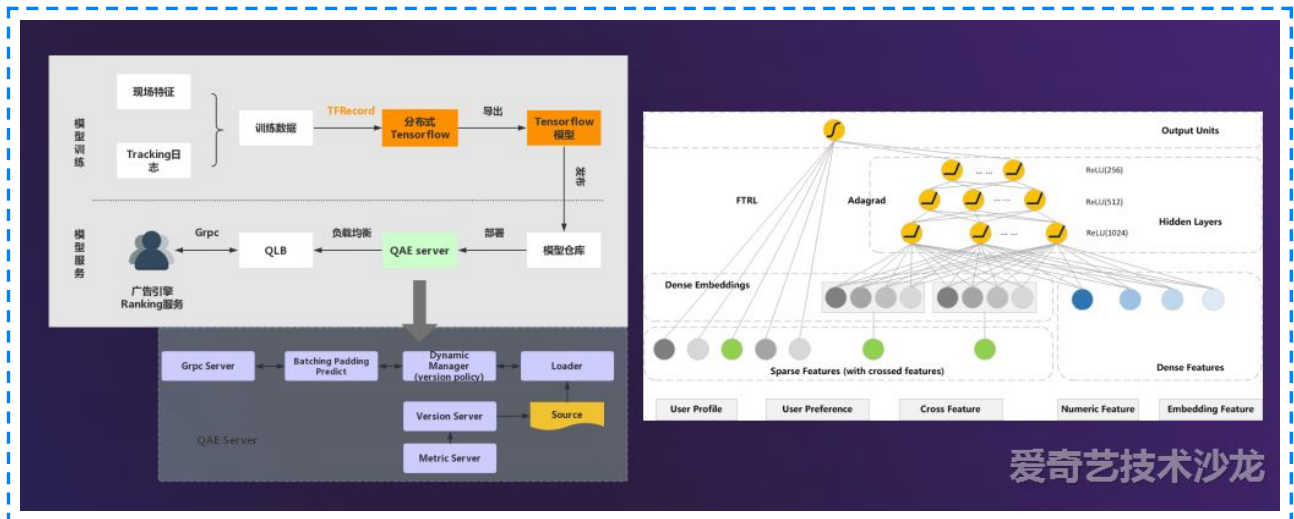
- 实时特征**主要考虑**场景特征**，即上下文特征，比如用户在看到广告之前看了哪些视频；**时间**信息也是非常重要的实时特征，比如周末和工作日分时段的点击转化率表现就有很大的差异；此外还有**反馈**特征，包括正反馈及负反馈，主要针对用户看过哪些行业的广告及后续转化行为。
- 短期特征**包含用户观影兴趣、搜索行为、社交行为和商业兴趣等。其中短期观影兴趣可能会发生变化，比如上图，如果有位女性用户怀孕了，她的观影兴趣可能会从长期的仙侠类视频短期切换到母婴类视频。
- 长期特征**有人口基础属性(性别年龄等)、观影兴趣、商业兴趣、广告基础属性(所处行业等)和素材质量等。这里的观影兴趣，依然举上图例子，这位女性用户可能喜欢参与仙侠类的讨论，喜欢出现在仙侠剧中的偶像等等，所以针对这位用户打得长期观影兴趣的标签是仙侠。

2. 在线学习



天下武功唯快不破，时效性对模型来说是非常重要的。当前线上使用的基于 FTRL 的 FM 模型，实时消费线上 Kafka 流的数据，分钟级进行模型更新。由于线上模型是分钟级更新，线上要有一套监控体系，一旦线上指标出现异动，比如上图举例的 AUC 低于离线分天训练的备用的模型，会进行线上模型与备用离线模型自动切换。

3. 深度学习



为了提高模型准确率和泛化能力，上线了深度学习，当前采用的是 Wide&Deep 的模型，并进行了离线模型调优和线上性能优化，以满足广告场景下对高 QPS 以及高稳定性需求。

4. 高维稀疏

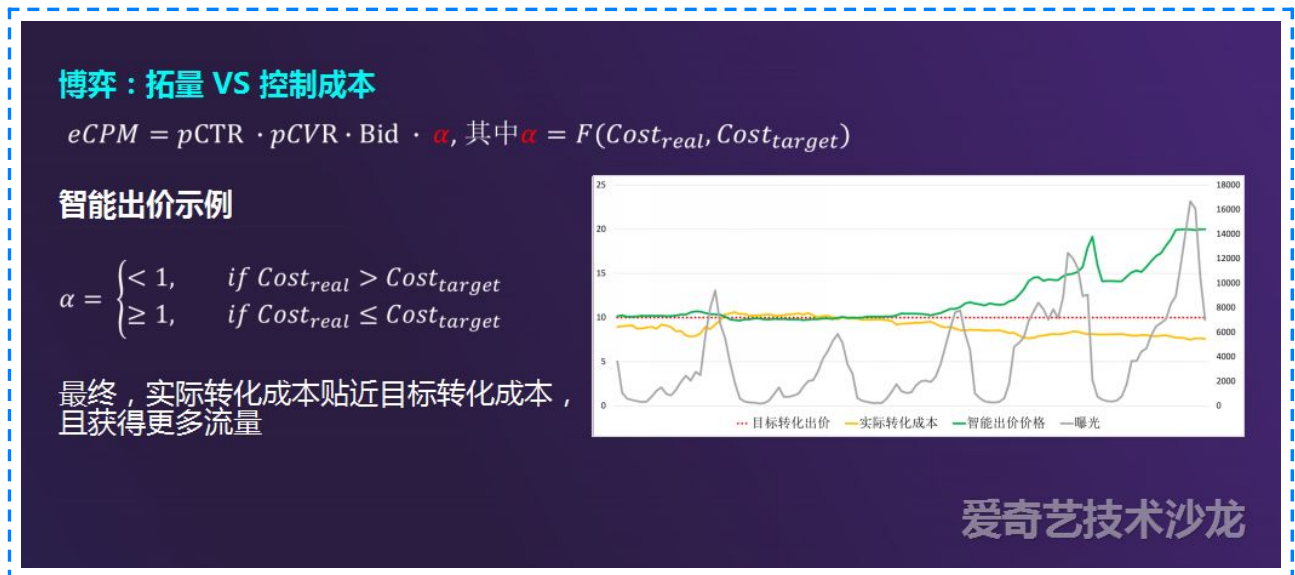
在转化率预估场景下，会遇到数据**高维稀疏问题**。举个例子，在做偏差校准的时候，会出现很多分桶里面一个正样本都没有，进而导致这些分桶最终的预估转化率结果为零。为解决这个问题，采用的方法是动态分桶大小，根据当前广告下历史正样本数量，动态决定分桶宽度以及分布。

同时，转化率预估场景有**滞后性问题**，例如以 APP 作为推广目标的广告主设置转化目标为付费，但从用户点击广告到最终付费可能会经历很长的时间。传统的做法是设立时间窗口对这条样本进行跟踪，窗口内完成转化目标即打上正样本标签，反之则当作负样本看待，但很明显这么做有很多问题。为解决这个问题采用了一种替代方案，当点击产生时，我们将样本当作负样本来看待，当转化行为完成后，这个样本就视为正样本，当然这里会进行适当的权重调整。

此外，还有**转化率阶跃的问题**，即转化数据集中到来，正样本突然增多，转化率预估可能从万分之几直接到百分之几。这样就会造成转化数据到来前广告放量不足，到来后爆量问题。为了解决这个问题引入非目标的转化数据作为正样本使用，比如转化目标是付费，我们可以引入下载、激活等也作为加权降权后的正样本来消费。

智能出价-博弈&共赢

1. 成本控制



智能出价的基本功能是进行成本控制，保证广告的实际成本在目标成本附近。并在此基础上尽可能多拓量。所以这里存在成本和放量的博弈。所使用的方法是在 eCPM 上乘一个智能出价因子，这个因子是实际成本与目标成本的函数输出，比如上图实际成本大于目标成本时，因子小于1，用更廉价的流量来拉低成本，反之大于1，提高竞争能力获取更多流量。

2. 流量优选

放量中: 以高出价获得了低质流量进而导致放量效果不佳
缩量中: 因低出价错过了高质流量进而导致成本控制效果不佳

$$eCPM = pCTR \cdot pCVR \cdot Bid \cdot g(\alpha, x)$$

其中 x 代表流量质量, $g(\cdot)$ 是关于 x 的单调递增函数

爱奇艺技术沙龙

上面介绍了基础版本的智能出价方案，但这样会产生一些问题，比如放量过程中的订单，可能以高价格获得了低质的流量而导致效果不佳，反之可能错过高质量流量。解决方法是做流量粒度下的智能出价，即考虑成本的同时也考虑流量质量，就是上图g函数，考虑成本参数的同时也加入流量质量参数。

DataFunTalk

分享嘉宾



王晖

爱奇艺 | 资深工程师

—END—

文章推荐: