

运营商数据在 DSP 广告中的使用方法。

海淘科技为你提供 DSP 广告最新资讯：**运营商数据在 DSP 广告中的使用方法。**

现在业界正在热烈讨论运营商如何去电信化、避免沦为管道商。运营商通过数据挖掘来开展增值业务的实践也早已开始。不过现在情况有了新的发展，美国的电信运营商的步子迈得更大了。据华尔街日报报道，国外运营商开始将自己手握的海量用户数据变现，将用户的位置、旅行、上网流量**惯等信息出售给包括广告公司在内的第三方。

那么在国内运营商数据如何使用才能发挥最大价值呢?本篇文章会介绍到

- 1.实时竞价算法(RTB)与 Ad Exchange/DSP 知识简介
- 2.DSP 中的实时竞价算法的主要功能及关键问题
- 3.运营商数据如何在算法各个环节上使用形成闭环的机器学习

一、实时竞价算法(RTB)与 Ad Exchange/DSP 知识简介

首先简单介绍下实时竞价的应用典型场景:



实时竞价会发生在 Ad Exchange、SSP、DSP 这些参与媒体流量售卖与购买的各个系统上(现在一个更广泛的概念叫做程序化购买/售卖)，本次交流主要关注在 DSP 侧的实时竞价的主要问题及我们的解决方法。



DSP 的最简模型如上图所描述。简单的说就是根据广告主的用户定向条件通过实时竞价的形式完成交易。

接下来就参考文献和个人的理解，对目前 DSP 市场发展的现状分析：DSP 作为一个产品形态，出现在中国已经近 3、4 年的时间了。随着大数据的火热，有数据资源的公司都开始想着数据变现的方法或者说商业模式，而基于 RTB 的广告是互联网数据变现的最直接有效的途径。目前国内的 DSP 厂商应该也在百家左右(也有大于 50 家的说法)，可是真正有规模有口碑的 DSP 依然是屈指可数。中等规模的 DSP 局限于自己的客户规模，目前也发展遇到了不可避免的瓶颈期。下图是目前国内程序化广告技术生态圈。



DSP 在这个市场中的占比也会不断的提高，这是市场趋势和技术进步的不可逆转的方向。谷歌，淘宝，百度，腾讯都拥有自己的 AdExchange 和 DSP，虽然目前的 AdExchange 和 DSP 真正盈利的并不多，但大家都没有放弃这个市场反而大张旗鼓的挺进这个市场，正是因为大家都看到了这个趋势。基于 RTB 的程序化购买和售卖已经浸入到媒体(Private/Public SSP)、交易平台(AdExchange)、广告主(Private/Public DSP)三个主体之中。同时也促进了第三方检测技术的厂商，如秒针和 AdMaster 等。

二、DSP 中的实时竞价算法的主要功能及关键问题

DSP 是对广告主的需求负责的系统，他的基本功能包括：人群/媒体/地域/时间等定向，曝光频次控制，价格优化，预算控制等等，最终目标是完成广告主的目标，包括：曝光，点击，注册，收藏，订单等等。通常的用户定向描述如下图所示：



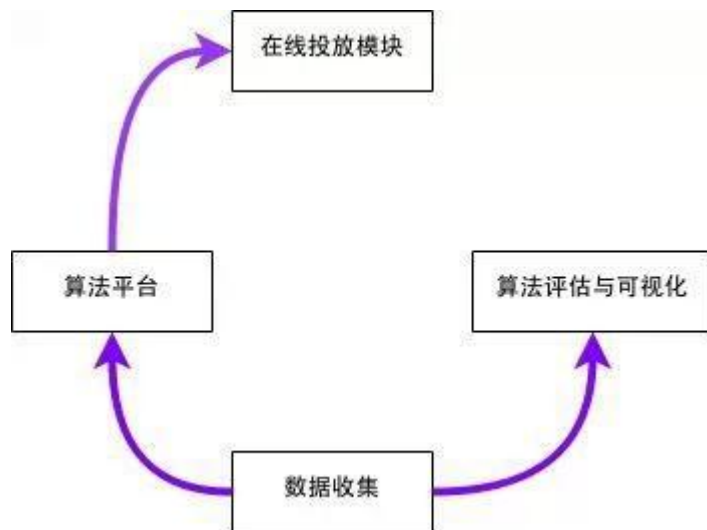
一个广告主入驻 DSP 需要销售合同的签订，运营策略的制定和实施，运营数据的支持与分析，运营日/周报，运营结项报告等等运营支持，当然背后也需要产品技术的响应支持与改造。下面分享一下我们团队对 DSP 系统的思维导图：



上述思维导图可能看不清楚，将算法系统放大如下：

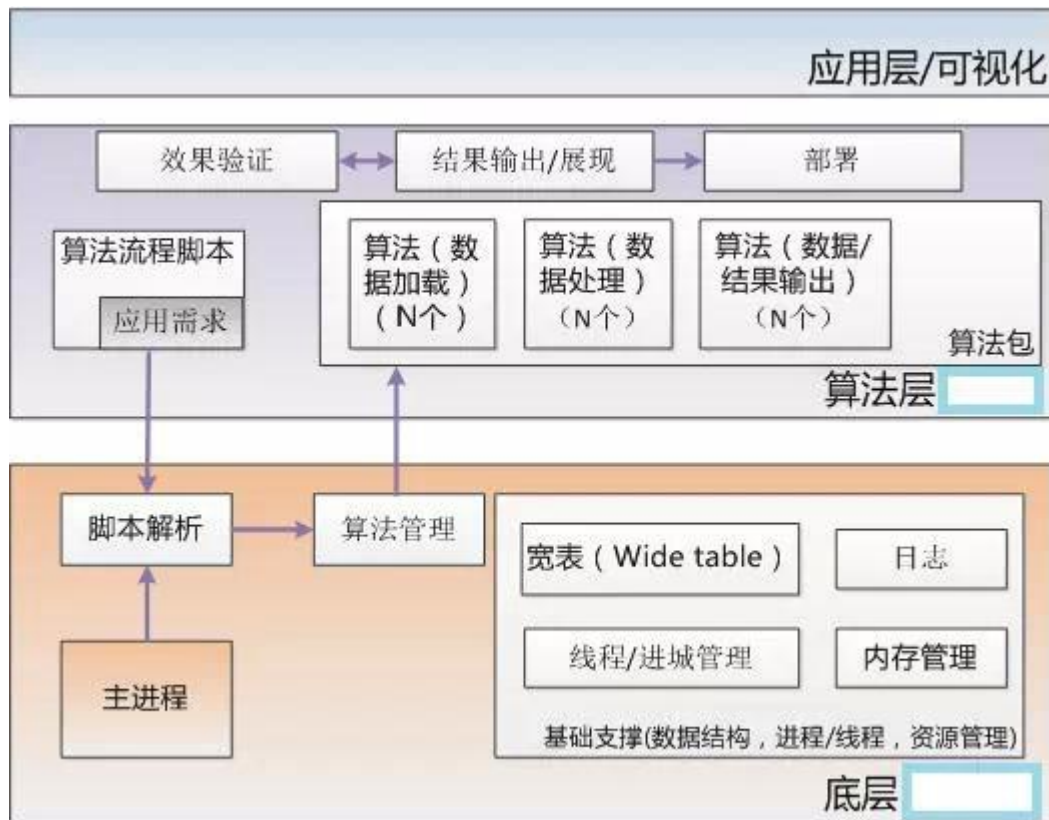


基于上述思维导图，形成我们的算法整体架构，方便大家理解算法整体架构简单描述如下：

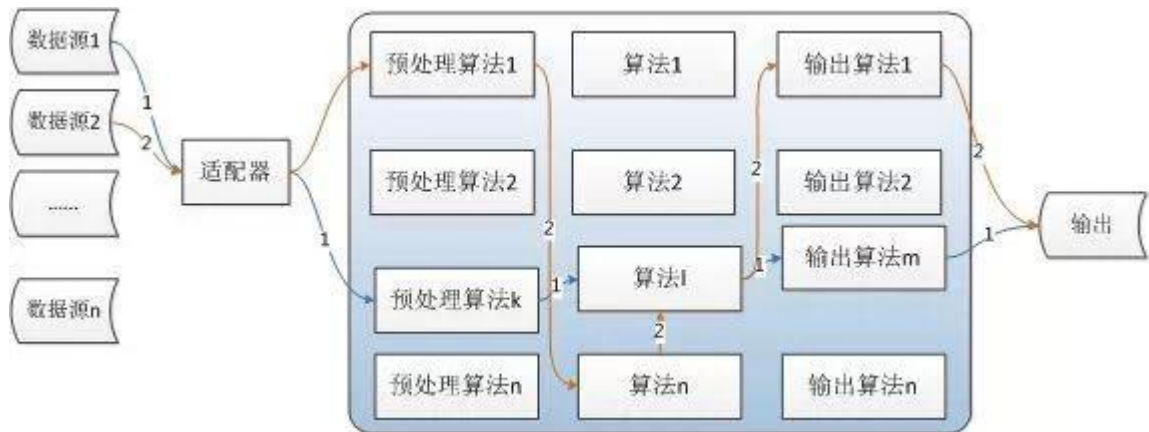


算法整体由四大部分构成：在线投放模块、算法平台、算法评估与可视化、数据收集模块。在线投放模块负责在线投放、算法与策略实施，并给每次投放打上算法标签方便跟踪分析。算法平台完成模型训练与离线评估优化，基于 Hadoop Spark MLlib 技术及自有算法包开发，模型以 PMML 标准语言描述，方便与在线投放模块交互。数据收集模块负责投放与效果收集。算法评估根据算法标签及数据收集，利用典型指标进行可视化评估。

上面介绍 4 个基本模块的功能，下面介绍下关键模块的架构方法。在线投放模块基本架构采用分层模块化的方式，将竞价流程按照漏斗模型分成几个典型层次，对流量进行逐步筛选，每层的算法模块可以做到热插拔和并行分流。算法平台的基本架构如下：



多算法组合示意图如下：



数据适配器，用来适应多变的业务数据。通过相对通用的算法包，结合适配器，适应多种形式的的数据源与算法业务需求。通过前面提到的适应热插拔和多算法并行的投放系统做到多算法并行投放。并结合算法评估和可视化系统做到多算法评估。多算法组合与评估，主要的评估手段为典型指标可视化支持下的 A/B 测试。

数据的闭环控制和自优化及适当的特殊广告主的策略支持架构是算法架构的另一个关键问题，需要做到平台整体优化与广告主优化的有机结合才能满足广告主千变万化的需求指标。随着市场的成熟，广告主越来越不满足于初期的曝光和点击，更多的关注进站后的有效行为。

如何做好 DSP 的算法是从本质上提升 DSP 效果的技术手段，也是困扰 DSP 厂商的核心技术问题。如何利用数据进行算法优化。就是利用这些数据处理协调好用户、媒体、广告三者的关系。在合适的环境给合适的人以合适的价格投放合适的广告。

要解决 DSP 的效果问题就需要对上面提到的三个关键因素及其相关关系进行彻底的分析 and 基于机器学习 (数据自循环) 的使用。

(1) 用户的分析: 用户的人口统计学属性、兴趣爱好、消费偏好; 广告主站内行为分析; 广告偏好 (如是否点击) 等。

(2) 媒体的分析: 广告位信息、内容 (特别是页面关键词)、分类

(3) 广告的分析: 广告主分类, 广告活动分类, 广告目标, 创意类型

(4) 环境的分析: 地域、时间、浏览器、设备、操作系统

上述四个方面及典型特征仅仅是举例, 不是所有。基于算上述四个方面特征的分析及相互的关联关系分析可以帮助大家更全面的分析流量特性及价值, 有利于在 DSP 购买过程中进行决策。

目前我们实现的算法模块及方法供大家参考:

点击率预测: (1)+(2)+(4)+(3) 核心算法: 统计/逻辑回归/其它

竞价价格预测: (2)+(4): 统计/对数正态分布/其它

基于目标的效果预测: (1)+(3)+(4)+(2): 稀疏事件/类目/统计

三、 运营商数据如何在算法各个环节上使用形成闭环的机器学习

1、 运营商数据的优缺点

运营商发展大数据具有其他行业无可比拟的优势, 主要体现在:

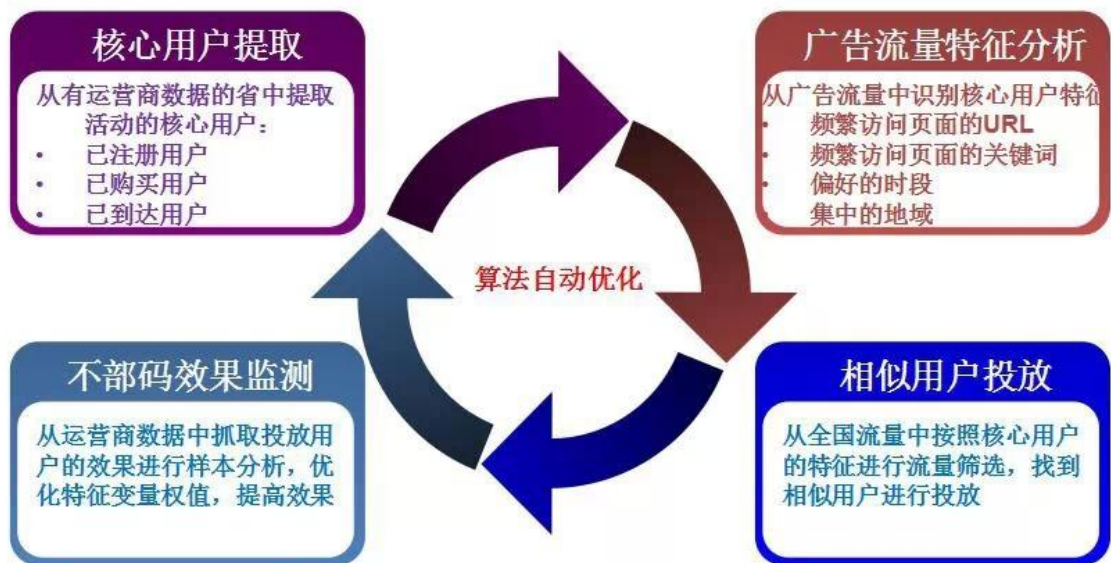
- 1) 运营商掌握数据全面充足。电信的领域之中, 数以亿计的通信用户基数保证了数据的海量和多元性;
- 2) 数据提供的可持续性。通信网络的时时存在为数据的持续和速度提供了保证;
- 3) 运营商对数据可以有效利用。运营商可以通过对海量数据的有效分析, 更加精准、更加高效地把握用户需求, 为广大用户和社会各界提供他们需要的产品和服务。

运营商数据也有其缺点:

- 1) 数据地域不同格式和字段差异
- 2) 敏感信息风险
- 3) 处理分析的难度较高

2、 运营商数据在 DSP 算法上的应用

运营商数据可以有效给 DSP 算法提供人群标签信息 (这个在周一的介绍中已经详细阐述, 在此不再详细阐述), 可以对 DSP 的投放效果进行收集和分析, 从而建立有效的机器学习机制, 有利于算法不断优化。目前我们已经形成了运营商数据利用的方法论, 跟大家分享如下:



运营商作为数据的管道,用户的行为数据均可以在其中采集和利用,如何有效的利用并闭环到 DSP 算法中,形成机器学习链路是 DSP 算法利用运营商数据必须解决的问题。经过 DMP 团队与 DSP 团队的紧密合作,在投放时植入投放特征代码,在分析数据时根据特征代码进行分析,结合统一用户识别技术,可是分析出用户经投放引领进站前、后的行为及动作,对投放的效果进行分析,并对投放人群进行优化,整体提高投放效果。

以上就是为您提供 [DSP 广告最新资讯](#), 同时提供 [百度 MIP](#) 资讯。本文下载, 点击: [运营商数据在 DSP 广告中的使用方法](#)。