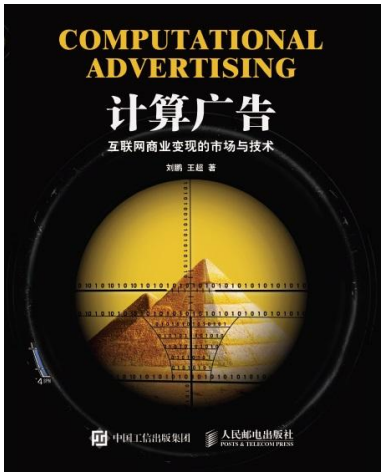
本讲座选自 2015 年 9 月 28 日刘鹏先生在清华大数据“技术·前沿”讲座上所做的题为《互联网变现与计算广告》的演讲。

互联网变现与计算广告

@北冥乘海生



“Comp_Ad”或搜“计算广告”

刘鹏：大家好，我是老的清华人，诸位都是 95 后了，你们出生的时候我是 95 年入学。清华电子系呆了十年，04 年年底博士毕业，毕业以后在 MSRA（音），我去的时候开复刚调到美国去，我师从布莱克宋。我接触广告是从 08 年底，现在京东的副总裁张晨老师刚从美国回来建雅虎的研究院北京分院，雅虎研究院我是北京最早的员工，雅虎当时是一个很有意思的公司，现在大家拿雅虎不当回事，其实雅虎当时还是很强的，它的市值也曾经超过一千亿美金。并且雅虎有一个特点，它的产品线什么都有，有新闻、门户、搜索、邮箱，当时在全球范围么还是比较领先的。它的变现的形态和广告的形态比其他的网站都丰富，那个时候我们接触到很多的有意思的产品，像搜索。日本的雅虎市场也是很大的，还做北美的广告，包括很多的广告形式都是从雅虎开始做的。

雅虎那个时候有两位科学家，我印象很深，一位叫安哥瑞现在在谷歌，是美国工程院院士，他希望把广告里面有意思的事情系统的计算整理成一个学科，他

跟另一位科学家普莱斯顿，是一个经济学家，他们两个在斯坦福开了一门研究生的课程，这个课程很遗憾的是大家看不到课程的全貌，因为我没有在网上找到全课的录像，或者是前几部分的 PPT。

后来大家从学术界和工业界对广告开始重视起来了，以前学术界不重视，但是工业界一直很重视。安哥瑞整理这个课以后，学术界、工业界都开始重视这个问题。后来安哥瑞想把这个东西系统性的总结和整理一下，但是他的工作繁忙，一直没有做。我的功力跟他差太远，但是我也想做一些分享。三年前我在伟伦楼开过一个系列的公开课，当时听的大部分是工业界的人，学生来的比较少。因为在校的同学你了解搜索，了解 socal GOLOBO（音）一旦进入互联网界，你会发现这绝对是互联网一个核心的业务，没有比这个事再重要了，因为大家挣钱全都是靠这个。清华的师弟宋波老师给我录了一些课程，放在网易课堂。

后来总结了这些东西，为什么总结这么长时间？因为互联网领域变化太快了，不要想有一个模特脱光了，你在写生画一天，其实他老在动的。就像广告行业这几年发生的变化太大了，不断想跟着工业界的节奏在走，但是发现新产品、新技术层出不穷。到今年之所以有一个总结，是因为移动的走向差不多有了一个模式。

大家来这个讲座我相信并不是冲着我落的，是冲着大数据这个题目来的，刚才主持人说我是大数据界的盆子，当然我根本不属于大数据界，我看他们这么火，我看着眼热，我经常发表一些大数据不同的观点。大家对大数据的认识和扎实的讨论是有点鱼龙混杂的，这是我真实的观点。我在里面写的文章都是玩笑之作，完全是调侃的口气，但是没有关系，只是表达我的一些观点。

戏说互联网思维之“三个不要”

- “不要钱”
  - 免费倾销加后向变现的商业模式
  - **所有能够传播信息的商品，售价都会趋向其边际成本**
- “不要脸”
  - 无底线迎合用户的产品与营销方式
- “不要命”
  - 用期权和价值观让程序猿在疯狂状态下全天候工作

今天借这个机会从我们做计算广告的角度谈一谈，对大数据大家像一个盲人摸象一样。我谈一谈我对数据的理解。所以我的题目叫做互联网变现和计算广告，谈他们两者之间的广告。

我们还是从一些不太严肃的风格开始，这是我在空号里面写的一个文章，二十一世纪还是什么一个杂志让我写一篇文章，谈谈互联网思维，我想了半天，因为我一直在互联网行业里面干，互联网思维是什么？我指的是中国的市场，中国

市场我总结了三个点：

第一，不要钱，和我今天这本书直接相关。互联网上最核心的一点商业模式的东西是免费倾销加后向变现的商业模式。嘀嘀、快的的模式不客气的讲就是倾销，但是不是每一个企业倾销完了都能够活下来，或者能够长大，企业在用倾销的方式获得了市场占有率以后，由于你推的是免费产品，怎么挣钱呢？就要大量用到后向变现的方式，后向变现就是把我免费产品获得无形资产变成钱的过程。我总结了三种资产：第一，流量，别人在用你的 APP 的时候你可以顺便在上面放一点东西，夹一点私活。流量通过广告变现大家都明白。第二，数据，大家都是奔着大数据来的，肯定对这一点很有兴趣，数据怎么变成钱呢？数据能不能挣钱，会有很多人问这个问题，我觉得这个问题特别可笑，数据不仅挣钱，而且是规模化的盈利，这件事情已经不是这两年才发生的事，这是十年前就已经发生的事情。为什么现在大家还在讨论数据能不能挣钱，这说明很多大数据领域的人并没有真正研究过去在互联网里对数据的使用方法和变现手段。我们觉得有一个规律：一切规模化、个性化传递信息的商品，它的售价都会趋向于边际成本，一个网站或者一个 APP 边际成本是多少，每多服务一个用户，他应该付出的额外成本是多少。边际成本应该是零或者是很小一个数，很自然的这些产品的定价都应该是免费的。其他的商品，比如说电视，乐视的电视是多少钱销售的？毛利为零，甚至是负毛利销售，他有非常明确的后向变现的手段，不要担心他挣不到钱，他只要能够做到一定规模，挣钱是板上钉钉的事情，他挣的方式是先进的方式，别的方式会被他的所打败。手机很明显也会趋向于零毛利的销售。有一些牌子的手机现在毛利已经很低了，甚至是负的，这都不奇怪。还有一些大家可以去探讨，比如说电影，我坚定的认为电影的票价绝对应该是零，这件事情什么时候会发生？以我最保守的判断，绝对不会超过十年。它的原理是，比如说最近有一个片子叫《港囧》，之前一部叫《泰囧》，它的票房非常好，等于卖给了三千万人，三千万人对于大众喜闻乐见的方式来说，太少了，如果我们用免费的方式把它变成三亿人看，后端产生的商业价值难道仅仅是十个亿吗？可是问题就来了，如果我们仅仅把前端的商业免费了，后端的变现我们不掌握，你的片子就白亏了，所以后端变现的体系是非常重要的。我本人也看过一些电影，包括植入的广告，包括各种形式，他们从植入广告这一点来说，他们的商业模式还属于比较低级的阶段，现在这种方式支撑不了把片子免费，还获得十亿以上的收入。这里面有很多利益相关方在里面。不要钱，如果你想知道我的书写什么，我希望大家了解什么，重点是了解这个东西，免费的流量和数据如何变成钱的，它涉及到很多复杂的产品技术。

第二，不要脸。现在大多数互联网产品的营销方式和产品点是无底线迎合用户的状态。特别是在面对比较年轻用户的时候，各个互联网公司在产品的文案上、产品的营销点上都是非常出格的，是跪舔用户的状态。

第三，不要命。在互联网上有一种工作方式叫九九六。9 点工作到 9 点，每周工作 6 天。这在很多创业公司和大一些的公司都是广泛存在的。为什么互联网的人能够这样疯狂的工作？关键的一点是全员持股，硅谷最核心的一个发明就是

告诉大家，这企业是你们都有份的，硅谷的全员持股是比所有的技术创新都重要。其他的技术创新是在这个基础上产生的，如果我每个月拿三千工资，你看那个公司能搞出什么。他是在这个激励下产生的。包括马老板说的，马老板把自己的股份给员工分了多少，他自己就剩下百分之七点几的股份，这个事情是大家拼命在阿里加班的前提，我觉得不是价值观。

与商业化相关的产品问题

- **商业模式探索**，例如：电影是一种边际成本很低，同时信息传播量又很大的典型商品。是否可能探索一种售价很低，而充分利用其信息传播能力的电影行业发行模式？
- **流量变现**，例如：互联网电视厂商除了销售收入，还可以获得用户流量。这些流量的性质如何，应如何变现？
- **数据变现**，例如：室内导航技术是近年来快速发展的新型互联网应用。这类产品会得到什么样有价值的数据资产，又应该采用哪种具体的商业产品来变现？
- **商业产品建设和运营**，例如：团购、游戏联运、返利购买这些推广模式与广告有什么内在联系？是否可以共用某些产品和技术平台？

我们重点看第一点，后向变现，或者叫商业化。

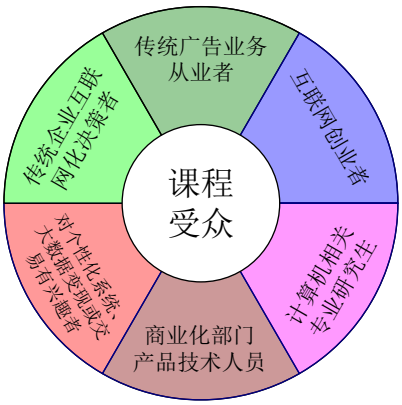
第一，商业模式探索。所有免费用户产品在做到一定量以后都会面临这样一个词：商业化。商业化是一个很大的领域，跟商业化相关的问题也很多。我这边举了一些例子，碰到这些问题你就要从商业化里面找答案，而不仅仅是要用用户的角度去找问题。

第二，流量变现。

第三，数据变现，我通过免费的用户产品，积累了一些用户行为或者其他用户相关的数据，这些数据怎么变成钱？近些年来大家发现数据变现的能力在某种意义上还要强过流量变现的能力。数据怎么变现？大家先不要去看大数据领域讲的东西，你先好好学习学习广告，因为数据的变现、数据的交易、数据隐私保护的边界在广告领域得到了充分的研究和工业界实战。你要不了解广告，你一定是从头走一遍弯路，这个弯路是非常多的。

第四，具体操作层面的东西，商业产品的建设和运营。比如说一个公司有广告，有游戏联运，返利购买，他们之间是不是有内在的联系？其实他们都是泛广告产品，他背后的商业逻辑基本是一致的。应该共用某些产品和技术平台去实现一个公司整体的商业化战略，这些大家在实际工作中才有感觉，前几个问题都是蛮有意思。我特别希望同学们如果在你的学习之余，除了了解一些用户产品，还能了解一些商业产品的思维、技术，对于你将来参加互联网公司的工作很有帮助。

哪些人需要了解商业化与计算广告？



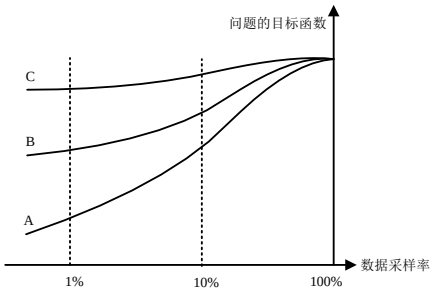
我的书的内容是基于我的公开课，后来在北大、北航都上过一次研究生课，总结出来的。对互联网创业者、对互联网行业的从业者，对计算机相关专业的研究生，我都希望他从这里得到一些东西，希望各位给我提一些建议。

从大数据说起，大数据这个词是一个咨询公司提出来的，这个词并不是来自于学术界。第二，也并不是来自于纯粹的工业界。这个词的立意非常好，让大家在大的场景下了解数据的价值和作用。由于它这样的起源，在中国现在的状态上，它与工业界实际发生的数据运用的现状以及学术界可落地的研究存在一定的距离，很多时候是概念到概念。所以我常常讲 BIG 是汉语的英译，是逼格的音译。我认为必须要找到一个落地的点来看看大数据到底做什么。

我自己对大数据的认识，我是从工业界来的，工业界对大数据最直观的认识是传统的工具用不了了，微博上有一些朋友来问我，我现在学大数据是不是应该学 SASS 这个软件，这让我觉得很难回答，我觉得跟那个没有关系，但是卖这个软件的人肯定跟我过不去。因为我们要了解大数据研究的是什么东西，传统的 IOE 的企业研究的是交易数据的加工和处理，交易数据的加工和处理是非常困难的，因为他要求正确率极高，一条都不能错，实时性要求极高，所以 IOE 整个这套系统就是 IBM、Oracle 和 EMC。你别以为现在拉一个互联网企业出来就能做系统，他们绝对是吹牛。可是互联网企业处理的大数据和 IOE 处理的交易数据有点不一样，我们关注的大数据是指行为数据，行为数据跟交易数据的区别，交易数据指业务实施过程中不得不计的数据，比如说存取款、利息，这些数据你不能不记，你不记你的业务没有办法开展，但是行为数据是可计可不计的数据，比如说网站的浏览日志。互联网企业一开始也不是想到要记这些数据，因为他的服务器自然而然的给它记下来了，后来他就想能不能给广告变现带来一点作用，于是他就开始挖掘这些数据。交易数据如果是 1，行为数据一般都在 100 以上。第二，它对一致性的要求是比较低的，网站的日志丢千分之一对大多数业务都没有关系。意味着原来 IOE 所有架构对于处理这种行为数据是不合适的，因为它太贵，我们要用一种更便捷、更低成本的方案来处理。所以工业界我们看到的变化是我们所

用到的工具完全的变掉了，去 IOE 化，阿里这么说他有他的技术，如果现在互联网企业一拥而上，把银行系统都换掉，那是灾难性的。可是原来 IOE 的你也不要轻易的说你们在做大数据，你们做的事情跟大数据严格来说也没有关系，你们还是在做传统交易数据的挖掘和整理。

大数据与计算广告的关系



A：典型大数据问题，解决问题的效果随着采样率降低显著降低，例如**计算广告**、个性化推荐等

C：一般数据分析，非大数据问题，解决问题的效果在采样率降低时变化很小，例如各种洞察、单维度统计等

B：过渡类型问题，解决问题的效果随采样率降低温和下降，例如文本主题分析等

这个图，A 曲线，我认为的大数据是什么样的，我如果数据可以采样，就不是大数据的问题，C 类的数据可以采样，比如说我要统计 360 在各个省的用户占比，显然是我先对用户数据采样，采样十万分之一。可是你现在碰到大多数的大数据都拿这样的案例在糊弄大家，他们把数据大，就当成大数据。这种问题的特点是稀疏的采样数据，结果不变，或者结果的基本不变。就不是大数据的问题。大数据应该是 A 种曲线。什么样的问题是典型的大数据问题？什么样的问题不能采样？所谓的个性化问题，广告是一个个性化问题，我们要对每一类用户描述他的行为特征和个性偏好。如果我采十亿人，这十亿人描述完了，我采样一百万人，所有的事情照做，你能影响的广告效果和空间的那部分人群就变成了一百万人，这个系统使得你的系统收益大幅度下降。比如说个性化推荐，依然不能采样。现在新的业务，个人征信业务，他也知道每一个人都做描述，所有的个性化问题基本上是大数据问题。我们也可以从另外一个角度理解大数据的应用。我个人是这样人为的，如果你的数据出来的结果是给人看的，不能成为大数据的问题，一定是要给机器看的，你要形成一个闭环的决策过程。

两类数据应用：洞察与自动化

- 洞察(Insight)
  - **全局或局部统计性的信息（统计数据）**
  - 例：财务报表、人口统计、百度迁徙地图等
  - 主要用于宏观决策支持，面向领导和运营人员
- 自动化(Automation)
  - **个体的行为特征信息（行为数据）**
  - 例：定向广告、个人信用、企业信息等
  - 主要用于微观业务实施，面向机器和销售人员
  - 无底线迎合用户的产品与营销方式

广告是大数据的最典型的应用。数据应用分成两类，一类是 Insight，洞察，比如说 360 对每个省的人口占比，这个结果打出来的是一张表，财务报表、人口统计、百度迁徙地图，这就叫洞察，洞察是整体上把握一些宏观规律，宏观的决策、运营人员和领导用的。这样的领域不能说没有大数据的问题，也有一些采样以后做不了的问题也存在，但是大多数问题跟大数据毫无关系。

另外一类应用叫 Automation，自动化，我输出的是个体的行为特征信息，如果我对十亿人分析完了，显然领导是不能看的，只有机器能看。在这种情况下数据的结果主要用于微观的数据实施，面向机器和销售人员。我个人觉得自动化的应用，大数据的成分要多很多，洞察的这类应用有很多跟大数据没有关系。我特别不希望大家被很多宣传带歪了，不能弄一张报表就叫大数据，那个叫商业分析。大数据简要说就是面向大规模的加工行为数据，并且把这个加工结果自动的反馈给机器做决策的应用。这是我的看法。肯定有很多人不认同，但是没有关系。

数据变现基本原理

数据怎么变成钱的。左边这个广告位投放的吉列剃须刀的广告，这个广告位卖一万块钱，是流量的价值，我每天来了十万人，这十万人看到这个广告，你就

得给我一万块钱。吉列是主要面对男性的广告主，我只给男性用户投吉列广告，省出来的用户都是女性用户，我找一个化妆品的广告投给女性用户，我找每一个广告主各收六千块钱。对媒体来说，投入产出比也提高了，我收到了一万两千块钱。我特别要强调，多出来的两千块钱是什么，这两千块钱就是数据变现的价值。你知道了每一个人是男是女，在原来一万块钱基础上可以凭空多挣两千块钱。仅仅知道一个性别就可以多挣两千，你要知道更多这个人的信息和购物偏好，你显然可以挣更多的钱，这些钱都是数据变现带来的钱。

广告对于数据变现和流量变现，你们在学校可能不太了解广告，但是它太重要了，我们从三个点说明它的重要性。首先整个互联网的意义来说，整个互联网行业的大半部分的收入是来自于广告，大概要到七成到八成左右。当然有人说互联网是不是没有别的挣钱方法了才用广告挣钱？这种说法是错误的。互联网公司做的产品好用，还是微软做的产品好用？免费产品一定做的比收费产品好用。因为在互联网公司里，用户产品的部门和商业产品的部门是分开的，管这个产品的老大根本不考虑挣钱的事，没有这个 KPI，没有把他所有的精力和能力解放出来，他可以全身心的服务用户。我的观点是：没有任何的收费产品现在还能做得过免费产品。你要想了解互联网，你如果不了解后向变现，不了解广告，你真的不可能彻底的了解广告，谷歌、脸书 90%以上来自于互联网广告，淘宝八成是来自于广告，腾讯一半来自于广告，腾讯游戏业务里面有很大一块是游戏联运业务，本质上仍然是 CPS 收费的广告业务，算上那一块应该有七成以上。百度冲 O2O 的业务，那是一个赔钱的业务，赔钱的业务把这个收入冲上去，很难说他将来怎么样，总之现在从利润环境来说，八成以上来自于广告是一点问题都没有，这是一个先进的商业模式，不是无可奈何的事。

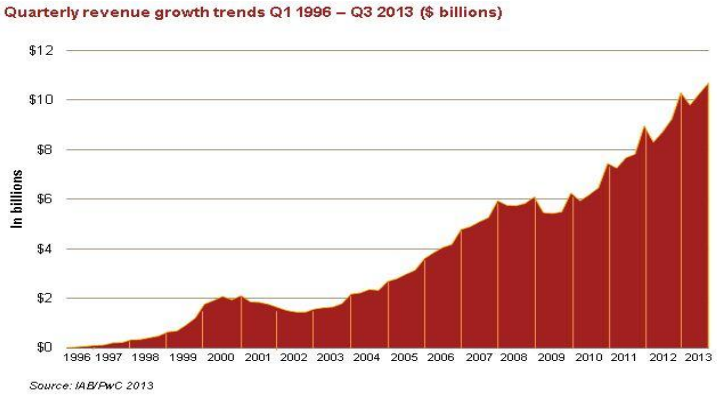
关于在线广告

- 在线广告支撑了整个互联网行业的大半壁江山。不了解互联网广告，就不可能深入了解互联网。
- 在线广告是迄今为止，大数据领域唯一形成规模化营收的应用。
- 在线广告是结合了计算技术、心理学、经济学、营销学等的综合应用。

大数据有很多应用，这两天很火那么规模化的应用我认为目前只有这样几个：个性化推荐是一个、计算广告是一个、个人征信正在试。普兰替尔（音）蛮独特的，但是它面向公众数据的大数据应用。但是其中形成规模化营收的行业只有广告。

广告很复杂，除了计算技术，还有经济学、社会学、心理学，都有非常具体的应用，非常具体的公式都要用上。

美国在线广告增长趋势



中国在线广告增长趋势

中美主要广告市场变化趋势

		2007	2008	2009	2010	2011	2012	2013
中国	网络广告	17	27	33	52	83	122	179
	电视广告	97	114	127	153	182	207	212
美国	网络广告	212	234	226	260	317	366	428
	电视广告	719	394	359	401	685	721	745
	报纸广告	486	344	246	228	207	194	180

规模化营收，在北美 2013 年在线广告总收入是四百亿美金，中国 2013 年达到了一千亿人民币，去年达到一千五百亿人民币。中国从 07 年到 2013 年在线广告涨了 10 倍，从 17 亿美金涨到 180 亿美金。对比的电视广告增长了一倍，全球没有增长这么快的电视市场了。美国从 07 到 13 年基本上没有涨，08、09 年的时候跌了很多，就是因为当时的经济危机，经济危机对整个广告市场的影响是巨大的。网络广告 07 年美国已经很成熟了，两百亿美金，但是它仍然增长了一倍多。报纸的数字惨不忍睹，中国的报纸可能跌下去的速度比美国还要快，我家附近的方圆一公里以内的几个报亭都没有了。特别是北京、上海，报纸跌的速度非常快。当然也不完全是，很多纸媒的老板跟我讨论，我们办电子版是不是就能解决问题呢？我个人认为是解决不了。这个图我希望告诉大家，在线广告是一个发展及其迅速的市场，它的季度复合增长率都达到两位数。并且这个增长速度现在没有变慢，而是在变快，因为移动互联带来了大量的新的机会。

传统广告主要做 Brand 为 Awareness，品牌广告，是为了带动长期的利润率和离线的转化率，他希望你记住这个品牌，将来选择它的可能性变大，承担的利润空间也会变大。可是互联网广告除了能做上面这种广告，互联网广告创造了一种崭新的市场---效果广告的市场。效果广告的市场有意思就在于，为什么互联网可以做效果？酷旁在线下发的效果是很低的，可是在线上发，数字广告可以很方便的对每个人投送不同的内容，短期有购买欲望的人一定是很少一部分人，数字化媒体特别适合做这个。你并没有看到在互联网广告快速增长的过程中，电视广告快速下降，其实没有这个其实，因为以谷歌、脸书为代表的互联网广告面对的是中小型的效果型的广告主，这部分的广告主传统电视广告对谷歌他们是不在意的。谷歌根本不屑于抢电视广告的生意，那些中小企业加起来比五百强的广告费多太多了。对销量比较在乎的情况下，长期的比如说到京东这种体量，他一定是效果和品牌要并重，只有品牌广告能拉动他的利润率，效果广告拉动不了企业的利润率。

说到计算，为什么上面这些事要用计算来解决呢？因为商业产品或者广告特别好的一点是，我可以用一个公式来表达我有话的东西，这一点比用户产品要简单很多。微信火了以后有很多分析师就来讨论，为什么微信比手机 QQ 好？但是这些讨论都是马后炮，或者并不能根据这些讨论重新造一些产品出来，因为用户是非理性的，我选择微信或者 QQ，有一些调研说 95 后更喜欢手机 QQ，这就证明在用户产品优化过程中很难找到一个明确的优化目标让他变得跟好。但是广告不一样，我们的优化目标很清楚。这个大括号里面有两项，一个是 R，一个是 Q，都是一个概念，没有任何数学成分在里面，R 是收入，你投一次广告出去挣了多少钱，Q 是成本，你得到这次展示的机会付了多少钱，这两个一减就是你的利润，你投广告的目的就是为了优化利润。前面那个求和，我优化的是一组广告展示上的总利润。广告跟个性化推荐最大的差别在，广告比个性化推荐复杂得多。最大的差别，广告主有预算，我今天最多投多少钱，还有一个是你今天至少要给我投多少，这使他的计算变得很复杂。R 有一个词叫做 eCPM——期望千次展示收益。

M 是一千次可能是几块到几十块钱，一次就是几厘，说起来很别扭。eCPM 是广告系统最想要优化的指标，提高 R，降低 Q。降低 Q 对于大多数的广告主来说不是一个核心任务。只有在 DSP 里面，Q 才是可以优化的，有一个出价策略的问题，大多数的产品主要是优化 R。广告的过程也很简单，但是也很重要，我们从广告的展示页，首先用户如果对他发生兴趣，发生一次点击，他在链接页上进行更复杂的操作，他如果想要这个东西，他会到转化页下单。点击的过程是发生在媒体上的，新浪上投的广告，点击是发生在新浪上，转化的过程是发生在广告主站内，点击和转化两个量发生在不同的媒体，这产生了一个有意思的分工，这一次点击到了广告主站内以后，他平均能够给广告主带来多少钱。这两个量的分解，决定了我们广告的很多有意思的付费模式。

品牌广告(Brand Awareness)

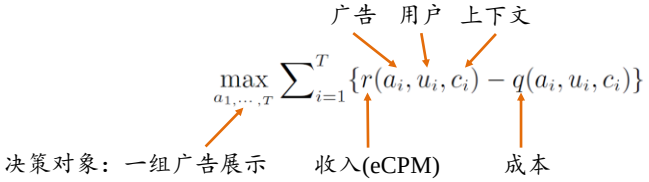
- 创造独特良好的品牌或产品形象，目的在于提升较长时期内的离线转化率

效果广告(Direct Response)

- 有短期内明确用户转化行为诉求的广告。用户转化行为例如：购买，注册，投票，捐款等.

计算广告核心挑战

- 计算广告的核心问题，是为一系列用户与环境的组合，找到最合适的广告投放策略以优化整体广告活动的利润。
- 优化问题描述：



广告收入的分解

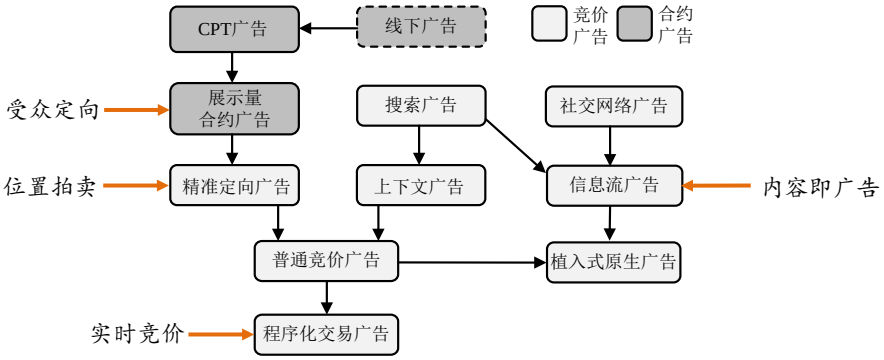
我重点跟大家讲讲广告产品的发展过程，让大家了解一下数据在广告业务里发展的核心动力作用。在广告行业里，我们的生产力是越来越多越精细的数据要用到广告产品的交易过程中。由于我们要用数据，我们在不断的变，才产生了现在非常复杂的产品形态。左上角几个灰色的我们叫做合约广告，合约广告是从线下广告直接演化而来的，线下的广告以杂志举例，杂志每期给你开一个位置，让你填上广告合同，这一期开你的，你给我多少钱。线上最早的时候也是这种方式，最早做这种广告的还是雅虎，当时最大的门户有一个叫做美国在线，美国在线跟雅虎是不一样的，美国在线当时是收费的，雅虎是免费的，不挣钱，雅虎就开出一个栏，当时是叫刊例价，你投在我这个位置上，投一天或者投几个小时多少钱，签一个合同，签完了我们就执行，这是最原始的方式。前面讲的数据变现的模式用不了，因为你把刊例的位置给一家放在那儿，他一定不是一个高效的模式，其实这种方式叫 CPT，按照时间来付费的广告模式。它主流存在的时间并不是很长，

很快进化到展示量合约的模式，展示量和约就是要用到我前面讲的数据量变现的模式。我把流量分成男女两部分分别收买。希望大家掌握一些商业产品设计和运营的思路。我一旦把修两分成男女在卖的时候就有一个问题了，你说你这里面一半是男的，一半是女的，这可不一定，比如说你是一个汽车网站，你九成都是男性，只有一成女性，你要告诉我，你应该给我投放多少次女性的广告。那个位置都给我，我可以雇公司来检查。但是你现在说女性给我，我没有办法检查，我只能要求你给我一个量的保证。我只能把展示量加在里面，这种叫展示量和约。在这儿产生了广告领域第一个里程碑式的技术和产生，我把人分成产品了，可以说也是一个根本性的变化，根本人群来售卖，售卖的标志已经变成人群了，不再是位置了。位置也只是一个载体。他让广告的售卖方式也发生变化，广告售卖方式要适用数据的使用，不得不发生变化。

这个变化进一步发展你会遇到问题，如果我希望用特别精细的数据来变现。我常举的例子是母婴人群，我们定义女性里面孩子在负一岁到正二岁之间的女性产品。他的变现价值高，因为这部分女性购物上呈现出全天侯且非理性的状态，我深有体会，因为我有两个孩子。我多卖 20%，这部分人群单价可能比正常人要高三四倍，我希望把他单拿出来卖。可是单拿出来有一个小问题，这部分人的量很少，有两重原因，第一重是确实量就不多，第二，我知道的不多。这部分人是母亲，我不一定知道，我知道是母亲的就那一点人，那一点人我单签一个合同，我会发现合同执行不了。因为那个量很小，就意味着不稳定。对原来的售卖构成挑战，雅虎现在的技术都解决不了这个问题，所以他的广告主一千到两千就上限不了了。还有一种方式是搜索，搜索的标的物是关键字，有的恨不得三个月来一次展示，你卖是不卖。

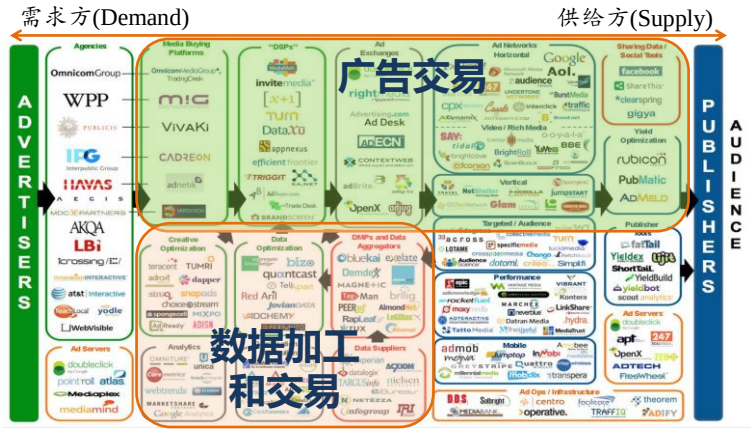
这两方面的要求需要有一种新的售卖模式，这种售卖模式就是我们真正广告产品上一个里程碑式变化：竞价广告。把定价权交给需求方，原来的定价权是供给方的。竞价的方式是这个东西你出多少钱，谁出的钱高，这个展示就给谁。数据交易将来关键的走势也在于定价权向需求方转移，我不跟你约定你拿不拿得到，你自己出价，拿到算你的。这样把整个市场盘活了，大量的中小企业主涌入到这个里面。

在线广告产品历程



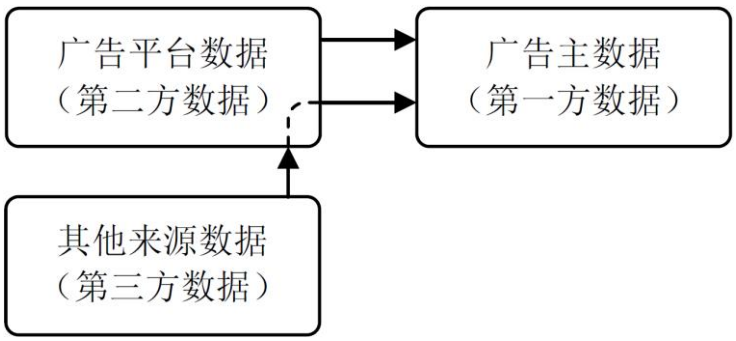
继续发展，又发展成现在的实时竞价广告，或者是程序化交易。这个词现在很火，它是最新的广告售卖模式，它的本质还是希望数据进入到市场里面，是第一方数据，第一方是指广告主。前面我们说的那么多数据，母婴也好，都是供给方给的定义，但是会有一些定义，比如说京东，我的流失用户这是我自己的定义，别人没有任何能力给我定义，因为你没有这个数据，前两个月来过京东，现在不来了，谷歌的数据再强你也不会知道。我希望用我的数据来影响我的营销。这种数据的价值是极高的，甚至远远超过第二方数据的价值。要想这种数据用起来，交易过程中，我没有办法预先开出来这样一个数据的展示让你来买。现在大家认为比较先进的合理的模式就是程序化的模式，我实时问你，我这里有一次展示的机会，在这个展示即将发生的那一刻，我把请求送到京东的服务器，问你一下，你要不要这次广告展示机会，你如果要，你自己定一个价格传给我，仍然是需求方定价。除了定价以外，把这个选择的机会也都交给了需求方。这盘活了很多东西，比如说今天的数据交易，如果没有需求方选择模式，数据交易量没有这么大，数据交易是程序交易规模化运转起来以后，才成为一个选择。

广告技术市场格局



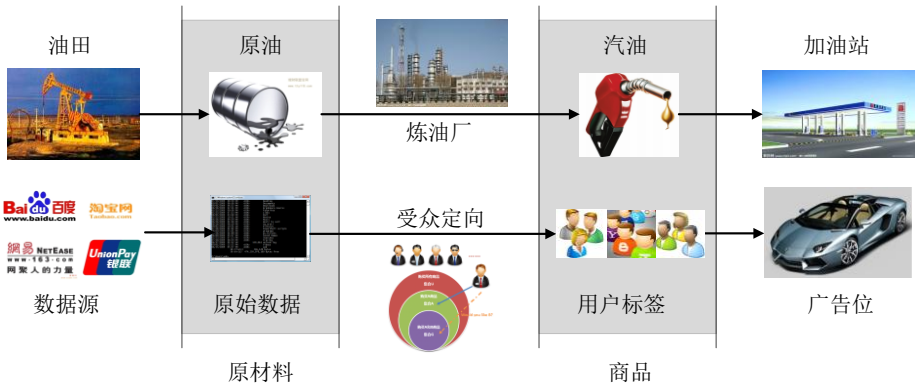
上面这个框是广告交易，下面这个框是数据加工和交易，但是下面这个是广告市场重要的支撑，我呼吁大家，如果你对数据感兴趣，对大数据的价值和交易感兴趣，广告里面的产品你是不能忽视的。因为这里面你确实已经做过很多东西了。它远远先进于其他行业所做的广告交易。

三方数据与数据交易



在这儿解释一下三方数据的概念，广告平台是第一方，广告主是第二方，其他的不跟广告关系的是第三方。

数据的价值与地位



广告系统是一个典型的个性化系统，它由一个在线的投放引擎，一个分布式

计算平台，分布式计算平台现在我们一般用的是 Hadoop，对于大量的海量的数据，我要对十亿的 Cookie，历史上三个月的数据做一次很浅的分析和挖掘，像这样大规模的数据，现在 Hadoop 仍然是唯一的选择，用 spark 也做不了，spark 适合中等建模。他们两个长期共存，各有各的优势。机器性能越来越好，spark 的能力越来越强，数据增长的速度比机器性能增长的速度还要快。流计算我们会用到，它的功能跟分布式计算平台是一样的，一个处理长时，一个处理短时的。

计算广告技术难点

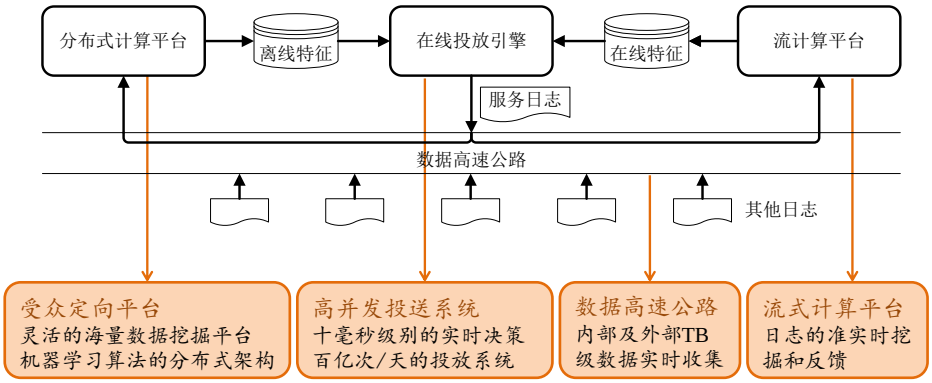
- 大规模(Scale)
  - 百万量级的页面，十亿量级的用户，需要被分析处理
  - 高并发在线投放系统(例:ADX每天处理百亿次广告交易)
  - Latency的严格要求(例:RTB要求竞价在100ms内返回)
- 动态性(Dynamics)
  - 用户的关注和购物兴趣非常快速地变化
- 丰富的查询信息(Rich query)
  - 需要把用户和上下文中多样的信号一起用于检索广告候选
- 探索与发现(Explore & exploit)
  - 用户反馈数据局限于在以往投放中出现过的(a, u, c)组合，需要主动探索未观察到的领域，以提高模型正确性

我们这个系统数据都是环形流动，我们尽量避免单点、高在线的同时读写。跟线上打交道的所有环节应该没有关系型数据库。你可以看出一个真正的大数据系统跟传统的商业智能和数据挖掘不一样的，我们尽量避免碰数据库，如果你线上系统发生了与数据库的数据交换，你一定不是一个自由体，一定是不太对的。他一定是更轻量级的，吞吐量更高的、容错量稍微高一点的系统来实现。

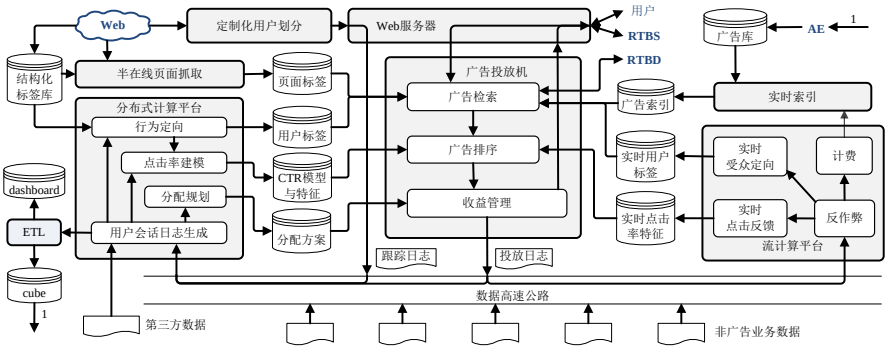
Web-scale技术问题比较

	搜索	搜索广告	显示广告	个性化推荐
主要准则	相关性	利润		用户兴趣
其他目标	垂直领域决定	质量、安全性		多样性、新鲜度
索引规模	十亿级	百万 千万级	百万级	百万级 亿级
个性化	较少的个性化需求		亿级规模用户上的个性化	
检索信号	较明确		较分散	
Downstream优化	不适用			适用

个性化系统一般框架



计算广告系统架构



数据交易是很有意思的一个问题，国内最近有两个数据交易所，一个是贵阳交易所、一个是长江交易所，我也关注了他们做的事情，我觉得很好，让大家认识到数据的价值，并且想办法用商业化的方式来运作数据，因为如果你不以商业化的方式来运作，这个数据很难用起来。可是我又看到他们在交易机制上的设计，或者他们对交易数据的理解，跟我认为的大家的状态有一点距离。数据交易应该是什么样的？它关键的问题和障碍都在哪？不是说现在广告市场对数据的认识就完善了，其实有很多问题还没有解决，这些问题是什么？我把总结成三个定律，也是我的看法。我跟一些业界的人交流过。

行为数据交易三定律

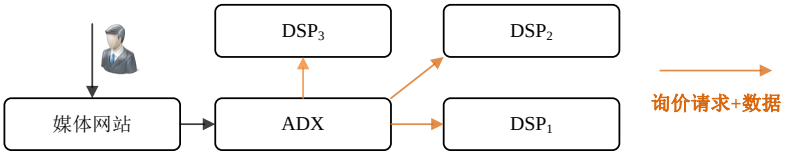
- 第一定律
  - 行为数据只能交易，不能交换或共享
- 第二定律
  - 只有按效果而非购买量付费，才有足够的需求
- 第三定律
  - 同一数据被越多人使用，价值就变得越低

关于第一定律：为什么数据不能共享？

- 疑问：数据交换似乎在发生啊？
  - 那往往是因为有更高层次的交换，即投资关系。
- 为什么大公司不把数据共享出来？
  - 你见过大公司把钱共享出来么？
  - 短时的贴补性共享是可行的
- 政府数据是可以共享的，这本质上是转移支付

第一定律，数据只能交易，不能共享。因为数据变成钱太容易了，你能够设想，现在有人在忽悠，百度拿数据拿出来大家共享一下，你还不如说让李彦宏把他帐户里面的钱打给你一部分。但是数据共享在有些层面是发生了，发生的情况有两种：一种是子公司和母公司之间，比如说搜狗跟腾讯是有数据共享，但是因为人家都是子公司和母公司，接近控股的状态。另外，政府的数据可以共享，政府的数据没有直接盈利、变现的需求，他的数据希望拿出来给大家用。可是我仍然觉得如果政府的数据简单拿出来共享也不见得是好的模式，还是要用商业化的手段做成转移支付的办法。总之我的第一个观点：数据是不能共享的，只有交易，交换也很难，一定是做价的交换，做价的交换本质上就是交易。

关于第二定律：数据交易该怎么做？



- 数据传输附着在实时竞价过程中，无额外开销
- 需求方可以自由地选择需要的部分人群数据，并且按照实际的广告展示付费

第二定律，数据交易该怎么做？我看过贵阳和长江做的交易，他们的交易有一个最大的问题，在广告数据交易里面碰到过这个问题，并且部分的解决了。数据交易必须要实现部分的交易才会有真正的市场，我知道全国每一个人的男女，打成一个包拿出去卖，可能有人买，但是买的人会非常少。我就在华东五省投广告，我买其他省的男女买了对我都是成本。在广告里的数据交易比这还进了一步，不仅仅是部分交易，而且是按效果交易。你在 Xchange 上买了一个广告位，我赢得了这次广告位我才交钱，我不赢得这次广告位我不交钱。这也是把定价权向需求方转移的过程。我认为在将来任何一个行业，如果能做到定价权向需求方转移，这个行业就有机会做大。按照部分数据交易、并且按照效果交易，这是我们在广告市场里摸索数据交易得到的共识。上海电信我打过一些交道，他卖了一年数据，我觉得他也卖不下去，他的意思是我这个数据很值钱，我打一个包给你，我只能用上一部分，其他的对我都没有用。也许今年我这个系统还在开发，今年数据我根本没有用，也交了一部分钱，这个事情表面上看起来上海电信占了便宜，但是他是吃了大亏，因为他没有真正把数据用起来，更谈不上将来用竞价的方式获得更高的收益。

关于第三定律：如何给数据定价？

- 市场化的定价方式是唯一的选择
- 目前数据的价值是被低估的
  - 上页的交易方式并未限制数据供给次数
  - 这间接地抬高了流量价格，而低估了数据价格
- 能否采用竞价的交易方式？
  - 不限量供应的商品，是无法竞价的
  - 数据的限量供应怎么做？

第三定律，怎么给数据定价？这个问题广告商也没有解决好，有解决的方案，但是解决的不好。有一个数据交易平台叫 BLU KIY，4 亿美金卖给了 oracle，它的数据量很大，但是它不怎么挣钱。后来仔细讨论，今年我深入的研究了这个事，我觉得数据的交易跟流量交易不一样，它反而有点像，比如说你知道一个人是男是女，这个信息你是可以卖给很多人的，你卖给一两个人、十个人，你会发现他不一样，你卖给十个人以后这个数据就贬值了。这块地，有的瓷砖下面有金子，有的瓷砖下面没有金子。有个人有一个藏宝图，这个藏宝图对他来说就是数据。但是我们每个人都知道这个藏宝图，会发生一个事情，大家都知道这块地上有金子，大家先来抢这块地，大家先把这个地价抬高。十个 DSP 都知道，我都出一个高价去买流量，数据就向流量价值发生了转移，这是一个理论。我个人的设想，将来数据交易应该是一种限量的，流量本身是限量的，一个流量就是一个人投，不可能三个广告叠在一起投。但是数据可以卖给很多人，卖的越多越不值钱。能不能把这个量限下来，这个人是母亲，一定时间段只让产生三次或者五次交易，这样的话有可能解决问题，并且这有一个巨大的好处，这样有可能让数据交易也变成一种竞价方式。这个母亲的信息我只给三个人用，我给哪三个人用，你们自己来竞价，最后排的比较高的三个人我给你们用。一旦数据交易能够变成竞价的方式，并且是在这么细的力度上竞价，前提是我们能够部分交易，整体交易竞价也没有意义。整体交易本身需求就很少，部分交易的基础上如果能做竞价，这个市场才能真正把它的市场打开。

数据定价和交易本身是特别有意思的问题，而且有可能激发一个巨大的市场，而且这些问题工业界都没有解决，将来大家如果从事大数据，这是很有意思的一个点。

大数据隐私问题严重么？

- 隐私安全基本原则
  - A29：欧盟负责隐私保护条例制定的委员会
  - A29原则
    - Personal Identifiable Information (PII) 不能使用
    - 用户可以要求系统停止记录和使用自己的行为数据
    - 不能长期保存和使用用户的行为数据
- Quasi-identifier与K-anonymity
  - Quasi-identifier: 朝阳区，35岁，在360上班
  - K-anonymity: 北京市，30-40岁，互联网行业

大数据的隐私问题，我发现没有人明白这个事，交易所还好，他卖的协会数据并不见得很多，隐私问题运营商提一个词：脱敏。脱敏就是我们这儿对应的第一条：PII，有些信息是你绝对不能用的，因为有一些信息你卖给他以后，他可以直接接触到客户，比如说电话，我直接打给你了，家庭住址、身份证号、姓名，

这种信息在我们的原则里面很早就有 A29 的原则，不能使用。很多做数据的人理解的脱敏就是把 PII 去掉，我今天要告诉大家的是，PII 去掉解决不了问题。有时候你看到一张表，你看到你们单位的工资表，假设会计把前面的电话、姓名都抹掉了，就剩下下面的几栏，但是有他的岁数、部门、家庭住址，在公司这个范围内，你对你熟的人，你看这几栏，一看就知道是谁，这种信息单拿一栏都没有办法定位一个人，但是一组放在一块，对熟人来说，他可以很清楚的定位一个用户，这种行为叫半定位。

大数据隐私问题比想象的更严重

- 稀疏行为数据的新挑战
  - 从一个人观影或购物记录，能否反推他是谁？
  - 实际案例：Netflix推荐大赛，有人从数据集里发现了自己的同事是同性恋
  - 理论研究：Robust De-anonymization of Large Sparse Datasets
- 深度个性化系统也有隐私安全风险！
  - 相关研究课题是差分隐私(Differential Privacy)
- 隐私是大数据头上的达摩克里斯之剑

隐私的顾虑在于熟人之间的隐私，不在于陌生人之间，我们最在乎的并不是有人把数据库黑了，把那个东西八毛钱一条往外卖，这件事不是我们今天最大的顾虑，最大的顾虑是对你有一些背景调查的人，在一些环境里能够把你的信息定位出来，从而了解你更多的隐私信息，这些事情是我们真正的顾虑。脱敏是不能解决这些问题的，而且这个问题是根本无法避免的。对互联网的用户来说，他的行为比前面会计那张表还要麻烦，他的行为是极为稀疏的，曾经有一个例子，前两年微博上传的一个例子，清华的一个同学是王珞丹的粉丝，他看王珞丹的微博，仔细一条一条分析，他就分析出王珞丹住哪，几栋楼，几号。有一个最生动的例子，Netflix 推荐大赛，这个大赛很有名，因为它的奖金很高，它出了一个事，有人对这个比赛感兴趣，在数据库里看，特别凑巧看到一条记录，他发现这条记录是他的一个同事的，因为那个太清楚了，这个人看过什么电影，同时给了多少分，你都有这种经历，你跟你们熟悉的人会经常讨论什么电影好看，评价如何。我们这个屋子和清华大学很难找到两个人看电影的记录是一样的。对于你熟悉的人，你一条一条记录扫出来，你一定可以把你的朋友对应出来，要不计代价的找到某一个人的隐私的数据，对这种场景，成本不是顾虑。当然这个案例他也是正好碰到了，他把那个数据拿给他的朋友看了。这个同事还有一些影片是没有跟他沟通过的，那些片子全部是同性恋影片。至少这个同事不希望自己看同性恋影片这个事被其他人知道，这是他的隐私范围内的，所以他很恼火，他就把 Netflix 告上了法庭，结论是没有办法解决，大家要把它上升到比较模糊的态度去解决，

这件事情并不是一定不能解决，也有数学家研究比较前沿的问题，怎么从数学的角度解决这个问题。在互联网的稀疏的数据下面，在熟人的前提条件下，隐私在行为数据里很容易被破解。他真正的风险在哪？比如说电信把他原始的数据，上海电信就是卖裸数据卖过一年。它的风险在于，假设有一个人不计代价的一条一条看，假如我跟上海某个副市长有仇，我把他的数据一定是能够找出来，找出来这个副市长有没有贪腐行为，不见得发现不了。

总之，告诉大家要有这个认识，隐私真正的顾虑是熟人之间的，不是陌生人之间的。关心你的人，这种关心有可能是负面的，有可能是正面的。这种人对你的影响最大，由于他可以不计成本，并且由于互联网的行为数据本身是极为稀疏的，基本上没有任何两个人一样，所以他的风险是很大的。而这种风险在今天来看是被低估了，因为没有出现哪个副市长因为这事被抓起来，将来他一旦出事，一定是有大事的。总之我的一个判断，隐私现在来看是大数据头上的一个达摩克里斯之剑。

大概告诉大家一个观点，你要想了解数据的变现和数据的交易，了解计算广告是一个必不可少的环节，甚至说是最重要的环节。因为所有的数据使用的历史和产品发展的历史在广告行业走过一条完整的弯路。你没有必要再走一遍，后面这半段涉及到数据本身，从我的经验来看，数据的变现和交易都是有市场基础的，并且有它的价值所在，但是数据交易本身有很多问题，有的是在广告市场里已经得到了验证和解决的，有的是我们在广告市场发现问题但没有解决的，还有一些是隐私问题。

今天借这个机会，我希望以计算广告作为一个引子，从这一点希望大家了解行为数据的使用、加工的过程，将来这一定也是我们大数据市场非常重要的一块。看大家还有什么问题。

提问：从您刚才说的整个过程来看，您是不是认为现在的广告模式已经到了比较成熟和适合商业化的阶段？您个人对将来新的广告模式可能的突破点在什么地方？

刘鹏：我那个产品图我有一半没有讲，现在的广告产品比较成熟，这句话可以认为是正确的。因为它从 98 年到现在，计算广告已经发展了快二十年的时间。现在大家认为对数据已经使用过了，隐私问题我们有很多顾虑。我们忽视了一个问题，我们用数据的理论是我们希望了解这个人历史上看过什么，对什么感兴趣，但是对用户现在的场景和情景的把握，在过去的广告产品里是不够的，这就涉及到我们讲的原生的概念和新的广告模式，这在移动上越来越重要。这个我今天没有时间讲，确实，原生是今后一个广告市场从产品发展的重要模式。把数据和场景结合起来使用。就现在利用数据的广告模式，从广告交易到数据交易本身，我觉得是比较稳定、比较成熟的。

提问：我有一些朋友也做过广告实时竞价的事情，听他们的意思，做这

个事情如何能够确认数据的真实性是一个很大的问题，是使用方在定价，但是他们很多人都无法监督广告发放者你到底给我发了多少，到底发给谁了。这个事情有没有进一步解决的可能性？因为您刚才说，数据能不能只卖给一部分人，如果这样做的话，就像电影一样人为的提高了这个价格，您最开始的逻辑，因为数据再卖一份的边际成本是零，如果我是有流量的人，我免费去卖，这样才能把我的流量价值抬起来，更符合您一开始说的逻辑的方式。

刘鹏：对于有流量的人来说可能就是这么想的，数据只要提高流动性，提高流量价值就可以。但是很多数据提供方并不是流量拥有方，这个市场有意思就在于，有广告需求方、广告供给方、数据提供方，这些人的利益出发点都是不一样的，都是博弈的，对于数据提供方来说，不是我们前面说的概念，因为我们前面说的概念有一个基础，我这个商品本身是能够规模化个性化传播信息的。数据本身已经不再是能继续传播信息的能力。所以我觉得数据提供方，数据跟流量的性质是不一样的，流量我可以搭别的东西，数据并不能再搭别的东西卖。前面实时竞价的问题，这跟需求方定价不仅不矛盾，而且恰恰是一致的，首先，展示量的问题是可以监测的，数展示量，数男女这种有确定标准的，人头属性流量，这都非常容易监测，而且也有第三方监测公司。其他的标签，我说这个人是一个体育爱好者，这种标签有一个特点，你不知道什么算体育爱好者，我去年打过一次羽毛球算不算，对于这种模糊的标签恰恰是要用需求方定价的方式来解决，你不用管我这个东西对还是错，你觉得他对你值多少钱，你觉得我这个东西质量差、掺水，你出低一点的价格，那个人质量好一点，我出高一点的价格。因为标准上是没有答案的，不存在哪些人都是体育爱好者，不如说每个人根据自己的需求来，根据你认为他的价值来定价。数据也是一样，供给方定价比流量还不靠谱。有的数据对有些人很值钱，对有些人不太值钱，如果都是供给方定价，这个市场很难发展起来。

提问：在效果广告很好的情况下，品牌广告主打品牌的方式该怎么选择？

刘鹏：这个问题我回答不了，上周我们还跟 4A 电通集团做过一个讨论，品牌广告在数字化营销面前应该如何制订他的 KPI，现在市场没有对他有深入的研究和了解。现在做的方式，有很多人生搬硬套效果广告的考核，最差的例子是汽车行业。汽车行业现在的状态是所有的广告公司都跟汽车广告主说我最后考核可以给你带销售线索，几个电话打过去，但是实际上品牌广告做不到这一点。除了一些垂直媒体，车友会有可能，普通的这样说基本上都是在骗人的，他们都是线下买来的一些培训好的人打电话。平面广告到底按照什么规律制订一个合理的 KPI 去优化？还是要回退到对品牌认知和品牌美誉度的提升。但是具体的数字化的衡量指标我现在真的是说不出来，我也特别希望以 4A 为代表的广告公司大家能真正的从品牌广告主的核心诉求去研究这个问题。

提问：你刚才提到原生广告和情境广告，现在从一个实际的用户来看，天猫的、京东的或者在手机上推送的广告，包括网盟的广告个性化具体的表现都不是令人满意的。从你来说，做个性化的分析瓶颈在目前这个阶段来看表现在哪几个方面？上下文场景数据、情绪数据、情感数据，怎么采集、怎么分析？对模型的训练和效果的提升哪个方面去发力提升效果比较明显一点？

刘鹏：个性化广告和原生广告是两个纬度的意思。原生，首先希望这个广告跟内容长的差不多，搜索里的广告最典型，搜索里的广告和内容长的差不多，很多人分不出来广告和内容。你在微博和 FACEBOOK 里看的广告也有这种特点。还有一种是情境广告，用户当前在干什么，如果这一点做到原生，我们叫做意图原生，搜索表现好在表现和搜索都是原生。过去所有的广告媒体是不参与的，媒体只是把代码放一个京东的代码，他就不管了，月底结账。淘宝的自然源处理能力去分析页面里的上下文，从而得知用户意图，这件事情靠谱不靠谱呢？基本上他能分析的东西很浅，深入的东西必须要媒体的参与。媒体怎么样重新参与到广告交易的过程中，提供有价值的上下文信息，如果这件事情能做好，真正符合你的预期的、在你任务里的广告就会发生。这是产品和运营体系上的问题，不见得完全是一个技术问题。搜索为什么能做到呢？因为搜索的用户意图就是他要搜索的东西。其他的网站里面用户意图也很明确，你通过自然语言处理是分析不出来的。我的看法是让媒体重新参与到广告的投放、决策过程中。但是目前有很多产品和运营商的障碍，不是一两天能发生。淘宝处理能力再强，也不可能投出符合你情境的广告，必须跟媒体想办法从数据和运营层面上有一些接口。

提问：我对数据交易比较感兴趣，您今天提的竞价是很新颖的一个观点。您能提供一些更细节的考虑分享一下吗？

刘鹏：考虑蛮简单的，我的出发点数据交易从 BLUKIY 的角度来看，是没有体现它的价值的。数据很大，但是它不挣钱。但是流量的价值往上走，竞价更激烈，所以我进一步考虑，数据的交易跟流量交易不同的特点，因为它可以不限量供应，流量天生就是限量供应。这里面存在机会。这些说法基本上是我个人的说法，不见得很成熟。可是我认为，竞价这个点是所有人都在往这个方向努力。现在这个时代任何一个行业要想有爆发式成长机会，一定要变成需求方定价和竞价的模式，供给方定价模式不可能爆发式成长。这是很开放的问题，你也可以提你的想法，没有什么标准答案。

竞价有很多问题，但是现在数据市场没有真正被激发出来，因为纯做数据变现挣不着什么钱。大家真正的热情没有被激发出来。总是要想一个办法把有数据的人的热情激发出来，让他真正挣钱，他才能够市场发展的快。至于你说的有很多问题，或者将来这个模式比较复杂，我倒不觉得这是障碍。商业产品的市

场就是复杂，而且越来越复杂，用户产品发展的规律是越来越简单，给懒人用，商业产品的规律就是越来越复杂，因为他的目标很清楚，就是优化利润。只要利润提升 1%，我的系统多复杂一倍都没有关系。

提问：我是来自华阳民众的，我也是这个行业里面从业者之一。我们现在也在跟 BAT 谈数据的设想，我感兴趣的是 360 这一块对数据变现和数据开放的思路是什么样的？

刘鹏：我们公司跟 BAT 是有点像，或者更保守，基本上是没有数据变现的需求。这个事情首先也变不了钱，你现在这种方式其实挣不了钱。首先 BAT 也不可能拿这个变现，他变现了也挣不了多少钱。公司的态度，不会说这是有一条产品线来做这个事情。

提问：阿里都是电商的数据，都是效果的数据，他自己本身的流量池也非常大，阿里妈妈的整合，整合流量的变现，包括对品牌广告主的开放。

刘鹏：他开放有一个前提，他必须是这些广告主的落地在淘内。

提问：至少数据变现的价值增值，数据竞价的问题，如果流量被竞价，数据的价值也就被竞价了，至少他自己本身淘内和体系内的流量的价值也会变大。在这里面加成的效果，他变现的体量也不会小，至少他现在一家一家的品牌广告主在谈，越来越多的广告主在天猫开体验点，或者有自己阵地的品牌商越来越多了，这个流量消化在自己体系内，对他是价值不菲的。

刘鹏：原来阿里几任负责 DSP 的人也都比较熟，因为各个部门的利益是有博弈的，未必数据部门的利益一定跟公司战略是一致的，我在里面看到很多的博弈。首先阿里现在用数据的方式还谈不到数据变现，更多还是用数据去提升他内部的投放。因为你在阿里投，阿里知道每一个 cookie 男女，你并没有图，跟你在百度选关键字是一个道理，你必须在他封闭的体系内消化掉，你不能把你男女的 cookie 带出来。这跟数据交易和变现是两回事，还是传统的封闭体系的话题。

提问：现在一有一个尝试，拿出电信的数据做变现的尝试。这个就不仅仅是在阿里体系内。这个事实拟稿怎么看？

刘鹏：我首先要提醒他们隐私，因为他们都不懂。他们老跟我们说脱敏，脱敏是石器时代的观念，不是互联网时代大家对隐私的认识。出了大事是他们电信领导都兜不住的事。电信数据所有互联网公司都没有他强，第二，跨屏的数据，

电信运营商对行为数据的掌控力在弱化。百度改 CBS 了，电信就拿不到百度的数据了，淘宝也要改 CBS 了，都改了之后，电信就拿不到了。但是这是迟早的事情。BAT 在高价值业务上 HP 改 CBS 的事情是迟早的事情，所以我不是特别看好他们在行为数据方面长期的价值，但是地理位置和跨屏是他们天然的优势。

提问：刚才您说到一个例子，有一个黄金埋在这儿十个人知道，它的价值变弱了，同样的数据对每个行业的人价值不一样。这个数据对这家公司产生先到先得的效益，对于不同的行业会产生不同的效益。所以我认为数据开放可能会不同行业的拉锯会产生各个行业的应用。

刘鹏：您说的是哪类的数据？如果是行业数据在广告行业也没有太多的东西可以参考，可以单独的摸索和探讨。

正因为数据对不同的使用方的作用和效果是不一样的，最终我坚信一定要走向需求方定价，供给方定价永远都解决不了这个问题。

提问：现在在各大应用市场上有专业的广告传播平台，他既可以给广告主发布，也可以一键转发，用多种传播的渠道平台一键托管的方式，这样的话对广告主也有预算，传播费用在平台上是透明的。对于用户来说，广告主可以把用户阅读自己定价，方便用户来传播，用户看到这个广告，觉得他值得传播，他就产品。

刘鹏：这种叫激励类的广告，积分墙之类的。这种广告传播的效果要打折扣，虽然它的传播量很大。比如说积分墙，你玩游戏，让你下一个，得多少分，但是你下了以后马上就得删了，这是一个很大的广告市场，但是它的价格比较低，积分墙是正常广告价格的五分之一以下，同样的一个安装，因为它的后期效果很大。一些大的手机厂商，苹果对这个打击很厉害，主要它影响榜单的排名。这类广告总体上是在走下坡路。因为它有点违反本质的规律，你是用激励的方法在传播。

提问：APP 上广告主投放广告，激励用户帮助他们传播。

刘鹏：我正常买一个 CPM 的价格是 10 块钱，投这种广告他们只能给两块钱。

提问：我对广告的监测和追踪很感兴趣，大的广告公司和网盟公司都有自己的监测中心，第三方独立的监测和追踪系统在这个行业里面发展的趋势怎么样？

刘鹏：监测和跟踪，看监测什么东西，广告是展示、点击、转化。点击不需要监测，转化是需要检索的，主要是移动应用下载，移动应用下载要溯源，他比较复杂，转化的监测也是由第三方公司提供。国际上都是按照一次点击多少钱。FLAS、MAT 等等几个公司在做这个。展示是在人家的网站，你不知道，你要委托一个第三方来做检测。展示付费你才需要检测，如果你按点击收费就不需要间。CPM 付费大部分是品牌广告，品牌广告才有监测的需求，大多数效果广告没有曝光监测的需求。这个市场你可以计算一下，假设品牌广告的预算一百亿，会有 1%的钱交给监测厂商，这个市场就是一亿，总体的市场不大。转化监测也不是很大，监测本身不是一个特别大的市场。

提问：比如说 You Tube 一上来放广告，然后给你一个按纽就可以关了，但是国内的广告他就一直放，一直放五分钟，这两种方式哪种是对的？

刘鹏：我当然认为 You Tube 的方式是对的，如果他放五秒你就关了，你对他产生的效果不大，他放五分钟对你的效果仍然不大，你可能去干别的了。国内的市场是劣币驱逐良币的市场。他由于对广告主，我承诺了多少量给你，我不得不去完成，为了充这个量，插进来一些低质量的量。广告主会进一步调低他的单价，我再参更多低质量的，变成了恶性循环。我卖一个 iPhone 你给我一百块钱，我跟你说它什么都能干，你就买了，我得了一百块钱。但是卖完了以后的售后服务不是我负责了。我个人倾向于 You tube 的方式，国外一般都是这种方式。

提问：注意到您说的数据交易，您提到了两个交易所，一个是贵阳，一个是长江，现在数据交易不挣钱，而且量也比较少，现在这种交易所有盈利的能力吗？对于交易所未来的发展方向应该是什么样的？

刘鹏：交易所的盈利能力，关于行为数据的交易，如果按照这种交易方式，就是不 work，按照贵阳的方法。如果是统计数据或者其他数据，那跟卖报告是一样的。统计各个县的什么情况。至于怎么发展，特别是行为数据的交易，不管是标签还是更细的数据，一定要结合在具体的数据应用上做交易才有可能快速的发展。这个广告市场做的很好，因为数据交易都附着在广告上，他没有独立的数据交易，Xchange，你买的数据是广告直接带过来的，而且所有的过程都是顺的，一定要附着在具体的应用上去设计交易模式。纯粹的数据交易很难解决部分交易的问题，很难解决按效果交易的问题。更谈不上竞价的方法。

提问：第一，个性化推荐和计算广告的区别点在哪里？第二，关系型数据库，来做这种广告投放平台，根据它的计算模式应该采用什么样的分布式计算系统？第三，这个计算平台，针对计算广告学这个特定的业务和状态，怎么样来

设计一个能够支持他这种有限的条件的模式下，他的整个过程追踪和快速计算、实时处理的平台架构，应该在哪些点上考虑的比较多一些？

刘鹏：第一，推荐和广告的区别，广告的核心引擎跟推荐是差不多的，广告最大的区别是多了一个预算和量的保证。推荐是很多很多的利益博弈，推荐是一个游戏，广告真的是市场，市场就涉及到机制，这是推荐里面没有的。如果仅仅是排序、检索算法，两者是非常相似的，但是宏观的设计和保证下的优化是有非常大的差别。

第二，我说的主要是线上不适合用关系型数据库，线下还是可以用的。你有一个 sever，sever 在做决策的时候不适合做计算，你算的空间很小。如果再有关系型数据库，基本上你的成本很高，而且效率一定不是最高的。关系型数据库就是准，一致性高，一致性高对广告来说无所谓，我有一千次展示，有一次算错了没有关系。

第三，这儿也说不清楚，你要有兴趣可以看看我书里讲的东西，基本上对框架有介绍，这个在业界来说相对成熟一点。

提问：刚才你提到现在有一个困难，媒体怎么样通过技术更好的结合做出更好的广告，究竟哪些方面有难点，我作为一个文科生，广告方面内部要做哪些改变，要从哪些角度切入？

刘鹏：比如说马蜂窝是一篇游记，游记下面你要放什么广告？如果你是编辑，你知道这个地方应该放酒店，如果你让自然语言引擎分析，他会分析出来一大堆乱七八糟的关键字。应该是由编辑告诉广告投放引擎，我这儿需要的是一个酒店，是什么地点的。把这些信息给他以后，他就可以很合适的把准确的东西返回过来。这都有一个自动化的过程，把页面从结构化的信息得出来告诉广告引擎，这是将来原生的困难点和复杂的地方。因为要结构化的告诉你我这个广告类型是一个酒店，地点是在什么地方。跟现在的交易模式差别很大，主要是产品运营上要经过很长时间的演变才能到那个地步。

提问：第一，搜索广告，你是怎么看待 RTB 广告的，在移动端它的未来是什么样的？

第二，我在各种妈妈上面看各种广告效果不是特别好，不太适合我，他们把好的地方给到自己的网站，把特别不好的地方给到联盟的广告。能不能做到大家把好的地方也留给我们。

第三，您说未来的方式是 CPS，CPC 和 CPM 的发展空间还有没有？未来十年之后是什么样的？

刘鹏：你自己对广告的属性，加上它做什么东西，你做了原生化的筛选，你也想让阿里妈妈给你出那个广告，但是他现在没有告诉你。阿里妈妈你人告诉他没有用，首先他的产品就不支持，你需要告诉他的是结构化的东西，你需要告诉他我这儿要放一个酒店，酒店需要在北京，需要是五星级的。现在阿里巴巴的系统根本就不支持你给的这个东西，他自己首先得变掉，有的网站先意识到，先愿意给他这个东西，这个需要相当长时间的磨合。你自己谈的广告主我相信在百度那儿也有，只是他没有给你显示出来，因为你说的这些信息你没有告诉他，他也没有让你告诉他，这中间是有鸿沟的。

第三个问题，我从来没有说 CPS 是方向，我认为 CPS 不是方向，这三种方式 CPC 应该是最合适的方式，具体原因书里面有详细的分析，CPM 也是合理的方向，CPS 只有在转化流程是一致的情况下是合理的。移动应用下载，转化的时候，国内不一样，国外都是去 Googleplay 都是一样的，淘宝也是一样的，其他的各自转化流程不同的情况下，做 CPS 是很难成规模化的。我给京东带一个用户，转化率是 3%，国美带一个客户，他转化率是万分之三，那是因为你的网站做的差。最后都变成畸形的东西，比如说淘宝代销，你卖一双鞋给我十块钱，我就放在淘宝上卖，我比你那儿便宜五块钱。买走都算是我的广告带来的量。所以 CPS 我认为它不是一个方向。移动应用下载因为这个市场太大了，所以它是一个特例，一般情况下我认为 CPS 不是发展的方向。应该还是以 CPC 为主。