

社会计算 第五讲

数据交易

赵明昊

2023年4月13日

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：什么数据不能共享

➤ 疑问：数据交换似乎在发生？

- 往往因为有更高层次的交换，即投资关系

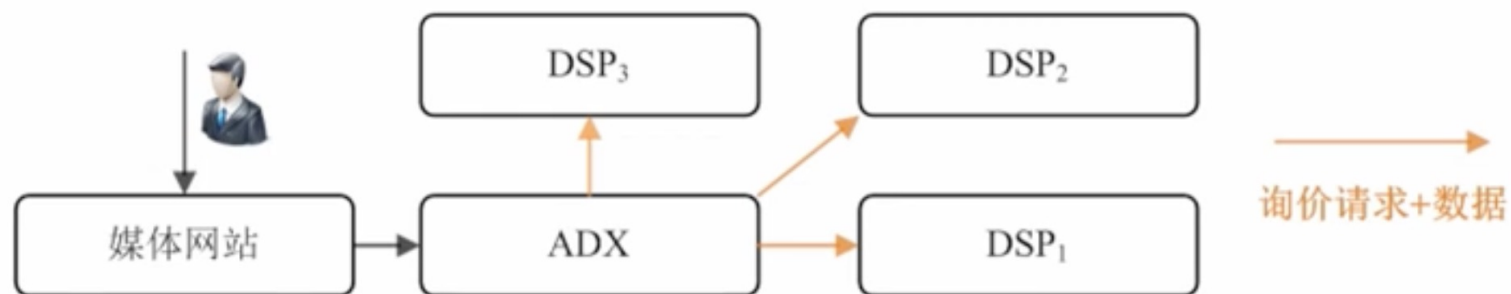
➤ 为什么大公司不把数据共享出来？

- 数据就是钱

➤ 政府数据是可以共享的，这本质上是转移支付

关于第二定律

- 第二定律：只有按效果而非购买量付费，才有足够的需求
- 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



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？

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