

从人群到个体的广告投放

2016-06-12

刘俊龙

概要

- 常用广告投放方法
- 常用广告投放方法的不足
- 可能的解决方案
- 我们的解决方案——从人群到个体的广告投放

常用广告投放方法

- 基于规则匹配
 - 基于城市、年龄段、职业范围、兴趣属性等的投放方法
- 基于特征工程的点击率预测
 - 基于用户画像、广告位上下文、目标广告属性，使用特征工程抽取特征，使用机器学习方法训练点击率模型
- 共性：
 - 都基于用户静态画像

常用投放方法的不足

- 1，数据稀疏时，用户画像不充分。
- 2，缺乏用户当前活动信息。
 - 在公司、在家、在运动场等不同场景下，用户有不同的活动信息，也有不同的行为习惯。
 - 时间和职业等标签能在较粗的粒度上反应当前活动
- 3，对用户兴趣转移不敏感
 - 用户兴趣受环境、周围群体多方面影响
 - 用户画像对转移存在一定的反应延时

可能的解决方案1

- 数据稀疏问题
 - 使用协同过滤、回归、分类等方法预测缺乏属性
 - 存在累积误差
- 使用用户公有数据项进行聚类，发现同质人群
 - 使用用户年龄、性别、城市等基础属性进行粗聚类
 - 使用用户实时位置信息进行聚类

可能的解决方案2

- 用户活动的获取问题
 - 根据时间和用户的职业进行粗略预测
 - 根据用户的实时位置进行匿名推断
 - 因为在同一位置内的人群基本做同一活动，所以可以将一定范围内的人群在同一时间段视作一个匿名的活动群体
 - 匿名推断，即使用位置进行活动id的标记，不需要明确知道活动内容

可能的解决方案3

- 兴趣转移问题
 - 对用户兴趣标签进行时间邻近程度加权
 - 在数据稀疏下，依然存在不准确问题
- 使用人群的兴趣补充用户的潜在兴趣
 - 人的兴趣转移受周围环境的影响，使用人群可以获取用户的潜在兴趣

总结分析

- 从以上3个问题的可能方案里，可以得出：
- 基于实时位置的人群信息可以缓解数据稀疏问题，可以匿名推断用户的当前活动，也可以获取用户的潜在兴趣
- 现在，我们投放系统存在的数据瓶颈：
 - 只有8千万用户的相对充分安装数据
 - 线上累积用户的app访问数量不到2个
 - 有10亿用户的年龄性别信息
- 现有数据还不能获取丰富的用户静态画像
- 所以我们可以尝试：从人群到个体的投放方案，补充用户的人群信息

从群体到个人的投放方案1

- 核心思路：
 - 使用用户的位置信息(ip)，构建不同粒度(ip段)的人群画像和广告点击行为信息。
 - 人群画像可以用来补充用户画像。
 - 人群画像和广告点击行为可以用来强化规则投放或概率模型里的特征工程。

从群体到个人的投放方案2

- 实施方法：
 - 使用线段树组织不同粒度的ip段，统计这个ip段上用户画像情况，对各个广告的点击率和下载率情况
- 使用方法：
 - 1，发现具有较高转换率的热点人群的ip段，固化成规则
 - 2，将人群画像和转换率作为特征输入机器学习模型

从群体到个人的投放方案3

- 可能存在的问题：
 - 细粒度的ip段可能存在数据稀疏问题。
- 解决方法：
 - 控制粒度；
 - 引入置信度；
 - 可以抽象为一个增强学习的问题，计算真实转换率的上下界， exploration-exploitation配合解决数据稀疏问题。（一个ip段上转换率的提升即Reward）

总结

- 常用广告投放方案有三个问题：数据稀疏时用户画像不充分；不能感知用户活动；对用户兴趣转移不敏感
- 针对这三个问题的分析，发现：利用用户的实时位置(ip)可以找到用户的人群属性、用户参与的人群活动、受人群影响的潜在兴趣。
- 最后，提出从人群到个体的广告投放方案：
 - 使用不同粒度的ip段发掘用户所属的人群画像和广告转换行为。
 - 既能丰富用户画像，也能强化规则投放和辅助概率预测。
 - 根据ip段的转换率上下界估算可以对广告投放的exploration提供参考。