

用户网络行为画像知识点纪要

原创 Thinkgamer 搜索与推荐Wiki 2020-06-05

收录于话题
#推荐相关笔记

32个



“

用户画像在推荐系统中扮演着重要的角色，这篇文章主要是阅读《用户网络行为画像》一书的读书摘要，该书较老，感觉更加适合产品经理或者不懂推荐的人看，如果读者本身就是一个推荐从业者的话，只需要看书的前半部分，对用户画像有个本质的理解即可，后半部分可忽略。

购买此书链接为：

广告

用户网络行为画像——大数据中的用户网络行为画像分析与内容推荐应用

作者：牛温佳 等著
当当

广告

用户网络行为画像 大数据中的用户网络行为画像分析与内容推荐应用

作者：牛温佳，刘吉强，石川，等
京东

当然也可以购买京东上的电子版，如果想要获取电子版，加我的微信号，备注【book】即可获取



用户描述分为三种情况

- 用户画像（User Portrait），更倾向于对同一类用户进行不同维度的刻画
- 用户角色（User Persona），更倾向于业务系统中不同用户的角色区分
- 用户属性（User Profile），更倾向于对用户属性层面的刻画和描述

一般将用户画像和用户角色合称为用户画像，而用户属性则是用户画像的子集。

用户画像分为：静态属性部分和 动态属性部分。因为用户画像包含动态变化部分，所以要设计合理有效的动态更新机制，从而精确的刻画用户。

用户画像的时效性和空间性

- 用户画像对于时效性非常敏感，某一时刻的用户画像对该时刻的推荐结果最为有效，距离时间越远，推荐结果的精确性越低，参考价值越差。
- 用户画像在不同的应用领域有不同的侧重点，比如营销领域的用户画像侧重于用户的消费习惯，而在视频推荐领域，用户画像则主要侧重于用户的观影喜好。

用户画像建模

- 用户画像建模往往就是用户信息的标签化
- 用户画像通常包含
 - 定量画像：主要包括用户基础变量，兴趣偏好等可量化的数据特征
 - 定性画像：主要包括用户的基本特征，行为刻画，兴趣建模和表征向量

用户画像的表现形式众多， 常见的有：

- 关键词法，即一组代表用户属性的特征词
- 评分矩阵法，以二维矩阵的行为表示用户，列表示特征项目，行列的交叉点为某用户对某特征项目的评分
- 向量空间表示法，给出一组关键词及对应的比重
- 本体表示法，以本体模型存储用户的属性和关系

用户画像的存储

- 关系型数据库（便于理解、通用的SQL语言使得操作库比较方便，维护成本低，如果达到了TB数量级，如果能够正确设计索引，并根据数据的时效性做进一步的分级缓存，减小缓存粒度，改进缓存机制也可以做到非常快）
- NoSQL数据库。键值（Key-value）存储数据集，比如Redis、Accumulo、BerkleyDB、LevelDB、Dynamo等；列存储数据库，主要应用分布式的海量数据存储，比如Sybase IQ、Hbase、BigTable等；文档型数据库（主要存储文档、半结构化的文档数据、以及特定格式的数据），比如CouchDB、MongoDB；图形数据库，比如Neo4J、InfoGrid、Infinite Graph
- 数据仓库，是一个面向主题的、集成的、随时间变化的数据集合，用以更好的支持企业或组织的决策分析处理。比较适合用户画像的存储、管理和分析

用户画像的查询

- 用户画像的用途决定了是以查询为主，因此强调需要有极高的并发查询性能，且同时对聚合性能要求很高。
- 避免查询重复，利用缓存技术

用户画像的更新

- 触发更新
- 定期更新



设计用户画像时，注意要有一个唯一的id表示用户，如果有多个，注意数据间的同步。

- 主流推荐算法分类
 - CF
 - 基于内容
 - 基于知识
 - 混合推荐
- 推荐系统的评测方法

- 离线评测：评分准确度-MAE（平均绝对误差），评分准确度-MSE、NMAE、RMSE，评分准确度-预测评分关联度（Pearson积距、Spearman相关方法、Kendall's Tau），排序准确度、平均准确度、NDCG（Normalized Discounted Cumulative Gain），分类准确度（准确率、召回率、F值、AUC曲线），覆盖率、多样性、新颖性
- 用户调查
- 在线评测：ABTest/点击率/转化率
- 常见的评测指标
 - 点击率&转化率
 - 用户满意度
 - 预测准确度
 - 覆盖率
 - 多样性
 - 新颖性
 - 适应性
 - 扩展性
- SVD算法
 - SVD的基本原理是将给定的矩阵 M 分解成3个矩阵的乘积形式，即 $M = U\Sigma V^T$
 - 分解出的矩阵 U 和 V 被分别称为左、右奇异矩阵，其本质是酉矩阵，即与他们各自的共轭转置矩阵相乘后会得到单位矩阵
 - 矩阵 Σ 为对角矩阵，其对角线上的值即为矩阵 M 的奇异值

混合推荐方法

- 整体式混合设计
 - 特征组合（从多种推荐数据源中获取特征，将这些特征组合起来放入单一的推荐算法中）
 - 特征补充（将一种推荐算法输出的特征信息作为另一种推荐算法的输入特征信息）
- 并行式混合设计
 - 加权式（对多种推荐算法的计算结果进行加权平均处理，产生出单一的推荐结果）
 - 混杂式（同时呈现出多种推荐算法的推荐结果）
 - 切换式（在不同的推荐算法中进行切换以适应当前的推荐情境）
- 流水线式混合设计
 - 层叠式（一种推荐算法对另一种推荐算法的推荐结果进行提炼）
 - 级联式（一种推荐算法学习到的模型作为另一种推荐算法的输入）

A/BTest（线上普遍使用），其宗旨为

- 同时运行两个或者两个以上方案
- 两个方案只有一个变量不变，其他条件均相同
- 有明确的评价指标用于评价两套方案的优劣
- 实验过程中，同一个用户从始至终都应该只接触一个方案



❤️ 原创不易，点个“**在看**”鼓励鼓励吧

收录于话题 #推荐相关笔记·32个

上一篇

你想知道的网易云音乐推荐架构解析，都在这里！

下一篇

机器学习在微博信息流推荐中的应用实践

阅读原文

喜欢此内容的人还喜欢

转化率预估中的贝叶斯平滑