

AI能力融合训练营-人工智能

第一天

引子

AI：让机器像人一样聪明

1. 推荐系统的发展

诞生

大数据加速推荐系统发展

概念：根据用户的偏好推荐最有可能感兴趣的内容的系统。

功能及应用

解决**信息过载**（数据爆发式增长）

涉及到两个方面：服务提供商和用户

是什么、干什么、用在哪

分类（3）

基于内容（content-based）的推荐系统——**特征信息**：关于物品的信息

- 主要基于物品相似性和用户画像。
- 存在问题：建议过于专门，新用户无法定义画像。

协同过滤（collaborative-filtering）的推荐系统——**用户-物品交互信息**：例如评分、购买数、点赞等

- 利用用户的交互来过滤感兴趣的物品，是基于历史交互信息预测新的交互。

混合推荐系统：避免单独使用一种信息引发的问题。

- 利用多种推荐算法配合，一共有七种不同的混合方案，根据方式不同可分为以下三种混合范式。

单体的混合范式

特征组合

特征增强

并行的混合范式

掺杂混合

加权()混合

分支(switching)混合

流水线混合范式

级联混合

云级别混合

难点问题 (4)

1. 新用户冷启动
2. 马太效应
3. 回音壁问题
4. 稀疏性问题
5. 灰羊效应
6. 投资组合效应
7. 稳定性/可塑性问题

2. 推荐系统的架构 (5)

召回系统

从海量的待推荐候选集中，选取待推荐列表。

排序系统

对待推荐列表的每个Item与User的关联程度进行排序。

数据层 (6)

用户数据、物品数据、用户行为数据（三种数据）

->逻辑层

召回算法、

->业务端

召回系统

主要解决信息过载和信息多样性差的问题

从还练信息中啥选出有价值的信息，缩小排序算法的搜索范围，解决信息过载。

将多路召回的数据进行融合，解决单路召回特征单一，信息多样性差的问题。

应用：主页推荐、相关推荐（二次消费）

排序系统

通过对生成的候选对象进行打分和排序

- 排序环节：**粗排**(规则排序层)、**精排**(学习排序层)、**重排**(后处理层) (7)

多目标排序：有多个目标函数，找到一种排序方法整体达优

- 目标偏差
- 物品偏差
- 用户偏差

常见模型

- 传统模型

Lr: LR是比较简单的线性模型，通过学习用户点击行为来构建CTR预估

GBDT+LR: 即利用GBDT构造的新特征来训练LR模型

- 深度学习模型

Wide&deep

3. 案例

手淘推荐系统

数据：基础数据：描述型数据，关系数据，图数据，行为序列

数据样本：Label、特征

优酷推荐系统

第二天
