

推荐系统（项亮）笔记

1 好的推荐系统

1.1 什么是推荐系统

推荐系统主要是解决用户和物品（信息）之间的关系。（信息数量>>用户数量）

面对信息过载问题的解决方法：

- 分类 --> 雅虎
- 搜索 --> Google
- 推荐 (是前面两种的互补)
 - 特点是不需要关键词，能自主挖掘用户的兴趣

现实中的推荐类型：

- 社会化推荐 Social recommendation：来自好友/朋友圈/小红书的推荐。
- 基于内容的推荐 Content-Based filtering：根据之前的搜索记录或者购买/使用记录，推荐相关物品。
- 基于协同过滤的推荐 Collaborative filtering：通过类似用户的数据，给出相应的推荐。

推荐本质上是将 用户和物品连起来。

1.2 个性化推荐系统的应用

推荐的应用集中在：

- 电子商务
- 视频网站
- 音乐平台
- 社交网络
- 阅读
- 基于位置的周边推荐
- 个性化邮件（筛选邮件重要性）
- 个性化广告

1.3 推荐系统的评测

1.3.1 推荐系统的实验方法

推荐系统的评测主要有三种实验方法：

1. 离线实验
2. 用户调查
3. 在线实验

离线实验将一部分用户的行为日志作为数据集，分成training set和test set。然后离线对模型进行测试。

优点：能离线测试大量模型。

缺点：无法获得商业化指标，比如点击率等等，并不能得到实时反馈。模型的评分高 并不等于 用户满意度高。

用户调查比如说让用户填写反馈，参与测试。

优点：能获得用户主观感受，不容易出现错误。

缺点：成本高，很难获得大量数据而造成结果统计意义上的不足。

在线实验将推荐系统上线做AB测试（分组测试）

优点：效果好，能得到各项指标的反馈。

缺点：周期长，难以确保其他所有环节都相等。

1.3.2 评测指标

1. 用户满意度
2. 预测准确度
3. 覆盖率
4. 多样性
5. 新颖性
6. 惊喜度
7. 信任度
8. 实时性
9. 健壮性
10. 商业目标

用户满意度

获得方法： 问卷调查/在线实验（对用户行为的统计）

预测准确度

获得方法： 离线数据集，建立模型预测用户在测试集上的行为。

- 评分预测：预测用户对物品的评分。

一般通过RMSE（Root mean square error）或者MAE（Mean absolute error）计算误差。RMSE对预测的不准确惩罚更大。

- TopN推荐：给用户一个个性化的列表推荐。

准确率通过 Precision/Recall 衡量。

$R(u)$: Recommendation list from training set.

$T(u)$: Recommendation list from test set.

$$\text{Recall} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|}$$

$$\text{Precision} = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|}$$

Recall主要是针对所有数据，衡量的是有多少被检索出来。

Precision主要是针对被检索出来的数据，衡量的是检索出来的有多少是正确的。

（通常会选取不同的N，计算一组Recall/Precision，并画出PR曲线）

覆盖率

描述一个推荐系统对物品长尾的发掘能力。

长尾：通常来说，80%的销量来自20%热门的品牌，而小众冷门的品牌销量很差，因此导致数据样本不足，模型拟合的效果也会很差。

最基本的覆盖率的计算公式： $\text{Coverage} = \frac{|\bigcup_{u \in U} R(u)|}{|I|}$ 能反映出有多少比例的商品被推荐了。

上述的计算覆盖率的方法并不能很好得体现商品推荐的分布，因此还可以用其他的指标来衡量：

- 信息熵

《信息量和概率的关系》

1. 概率越小的事件, 信息量越大.
2. 概率越大的事件, 信息量越小.
3. 多个事件同时发生的概率 是事件概率相乘, 而多个事件的信息量则是相加.

因此,用对数的形式来表示:

$$I(x) = -\log_2 P(x)$$

- 基尼系数（Gini Index）

