# 推荐系统

推荐系统是自动联系用户和物品的一种工具，它能够在信息过载的环境中帮助用户发现他们感兴趣的信心，也能将信息推送给它们感兴趣的用户。

# 推荐系统应用

pass

# 推荐系统评测

好的推荐系统不仅仅能够准确预测用户的行为，而且能够扩展用户的事业，帮助用户发现那些他们可能会感兴趣，但却不那么容易发现的东西。

## 实验方法

1. 离线实验
   1.1 获取用户历史行为
   1.2 通过历史行为建模
   1.3 对模型进行评测
2. 用户调查

- 优点：可以获取一些离线实验无法得到的评测信息，如用户的惊喜度。
- 缺点：用户调查成本高，用户需要花大量时间完成一个任务，并且用户调查样本少时不能保证结果的统计意义。

3. 在线实验

ABTest的方式对比不同的推荐算法。

## 评测指标

1. 用户满意度

- 通过购买率度量用户的满意度
- 通过满意和不满意的反馈按钮
- 更一般情况，用点击率、用户停留时间和转化率等指标度量用户的满意度

2. 预测准确度

评分预测：
一般通过均方根误差(RMSE)和平均绝对误差(MAE)计算。

$$RMSE = \sqrt{\frac{\sum_{u,i \in T}(r_{ui} - \hat{r}_{ui})^2}{|T|}}$$

MAE采用绝对值计算预测误差：

$$MAE = \frac{\sum_{u,i \in T}(r_{ui} - \hat{r}_{ui})^2}{|T|}$$

其中$r_{ui}$是用户u对物品i的实际评分，而$\hat{r}_{ui}$是推荐算法给出的预测评分。
TopN推荐：
TopN推荐的预测一般通过准确率(precision)/召回率(recall)度量。R(u)是根据用户在训练集上的行为给用户作出的推荐按列表，而T(u)是用户在测试集上的行为列表。

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

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

有时候为了全面评测TopN推荐的准确率和召回率，一般会选取不同的推荐列表长度N，计算一组准确率和召回率，然后画出准确率和召回率的曲线。

3. 覆盖率

覆盖率描述一个推荐系统对物品长尾的发掘能力。最简单的定义为推荐系统能够推荐出额物品占总物品集合的比例。假设用户集合为U，推荐系统给每个用户推荐一个长度为N的物品列表R(u)。

$$Coverage = \frac{|\bigcup_{u \in U} R(u)|}{|I|}$$

为了更细致的描述推荐系统发掘长尾的能力，需要统计推荐列表中不同物品出现次数的分布，如果所有物品都出现在推荐系统中，并且出现的次数差不多，那么推荐系统发现长尾的能力就越好。
另外两种定义覆盖率的方法：

- 信息熵

$$H = -\sum_{i=1}^{n} p(i) \log p(i)$$

- Gini系数

$$G = \frac{1}{n-1} \sum_{j=1}^{n} (2j - n - 1) p(i_j)$$

4. 多样性

多样性描述了推荐列表中物品两两之间的不相似性。因此，多样性和相似性是对应的。$s(i,j) \in [0,1]$定义了物品i和j之间的相似度，那么用户u的推荐列表R(u)的多样性定义为:

$$Diversity(R(u)) = 1 - \frac{\sum_{i,j \in R(u), i \neq j} s(i,j)}{\frac{1}{2}|R(u)|\left(|R(u)| - 1\right)}$$

所有用户推荐列表的平均值：

$$Diversity = \frac{1}{|U|} \sum_{u \in U} Diversity(R(u))$$

关于推荐系统多样性最好达到什么程度，举例说明。假设用户80%的时间看动作片，20%的时间看动画片。4种不同的推荐列表：A列表中有10部动作片，没有动画片；B列表中10动画，0动作；C列表8动作，2动画；D列表5动画，5动作。这个例子中，一般认为C列表是最好的，具有一定的多样性，又考虑到了用户的主要兴趣。

5. 新颖性

评测新颖度的最简单的方法是利用推荐结果的平均流行度，因为越不热门的物品越有可能让用户觉得新颖。

6. 总结

在给定覆盖率、多样性、新颖性等限制条件下，尽可能优化预测准确度。

max 预测准确度
覆盖率>A
多样性>B
新颖性>C

## 评测维度

- 用户维度：主要包括用户的人口统计学信息、活跃度以及是不是新用户等
- 物品维度：物品的属性、流行度平均分以及是不是新加入的物品等
- 时间维度：包括季节，是否为工作日，白天还是晚上等。
  如果能够在推荐系统的评测报告中包含不同维度下的系统评测指标，能帮我们找到一个看上去比较弱的算法的优势，发现一个看上去比较强的算法的缺点。