

评价实验方法

1. 离线实验
2. 用户调查 (user study)
3. 在线实验 (online experiment)

评价指标

1. 预测的准确度

其中TopN推荐：网站提供推荐服务时候，一般给用户一个个性化的推荐列表，这种推荐叫做TopN推荐
TopN推荐的预测准确率，一般通过2个指标

准确率 (precision)

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

召回率 (recall)

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

$R(u)$ 是根据用户在训练集上的行为给用户做出的推荐列表， $T(u)$ 是用户在测试集上的行为列表。

TopN推荐更符合实际的应用需求，比如预测用户是否会看一部电影，比预测用户看了电影之后会给它什么评分更重要。

覆盖率：是描述一个推荐系统的推荐物品的长尾的发掘的能力，推荐系统推荐出来物品占总物品的比例。

覆盖率是内容提供者关心的指标，覆盖率为100%的推荐系统可以将每个物品都推荐给至少一个用户。

除了推荐物品的占比，还可以通过研究物品在推荐列表中出现的次数分布，更好的描述推荐系统的挖掘长尾的能力。

如果分布比较平，说明推荐系统的覆盖率很高；如果分布陡峭，说明分布系统的覆盖率较低。