**1 评测指标**

评测推荐系统的指标可以考虑以下几个方面：

**1.1 用户满意度**

用户作为推进系统的主要参与者，其满意度是评测系统的最重要指标。满意度可以通过做用户调查或线上实验获得。在在线系统中，一般通过对用户行为的统计得到，例如点击率、用户停留时间和转化率等指标度量用户的满意度。

**1.2 预测精确度precision**

预测准确度度量一个推荐系统或者推荐算法预测用户行为的能力。这个指标是最重要的离线评测指标。由于离线数据可计算，绝大部分科研人员都在讨论这个指标。

评分预测问题一般使用RMSE、MAE等，TopN预测问题一般使用Recall、Precision等。

其实目前国内很多地方和资料混淆了两个指标的叫法，把准确度对应英文precision指标。不过尽量还是用英文比较好。

准确度Accuracy = (TP + TN) / (TP + FP + TN + FN)

精确度Precision=TP/(TP+FP)

**1.3 覆盖率coverage**

覆盖率描述一个推荐系统对物品长尾的发掘能力。覆盖率有很多定义方法，最简单的计算就是推荐列表中的物品数量，除以所有的物品数量。

在信息论和经济学中有两个著名的指标用来定义覆盖率，一个是信息熵，一个是基尼系数。具体公式和介绍可以google。

ps：长尾在推荐系统中是个常见的名词。举个例子帮助大家理解，在商店里，由于货架和场地有限，摆在最显眼的地方的物品通常是出名的、热门的，从而销量也是最好的。很多不出名或者小知名度的商品由于在货架角落或者根本上不了货架，这些商品销量很差。在互联网时代，这一现象会被打破。电子商城拥有几乎无限长的“货架”，它可以为用户展现很多满足他小众需求的商品，这样总的销量加起来将远远超过之前的模式。

Google是一个最典型的“长尾”公司，其成长历程就是把广告商和出版商的“长尾”商业化的过程。数以百万计的小企业和个人，此前他们从未打过广告，或从没大规模地打过广告。他们小得让广告商不屑一顾，甚至连他们自己都不曾想过可以打广告。但Google的AdSense把广告这一门槛降下来了：广告不再高不可攀，它是自助的，价廉的，谁都可以做的；另一方面，对成千上万的Blog站点和小规模的商业网站来说，在自己的站点放上广告已成举手之劳。Google目前有一半的生意来自这些小网站而不是搜索结果中放置的广告。数以百万计的中小企业代表了一个巨大的长尾广告市场。这条长尾能有多长，恐怕谁也无法预知。无数的小数积累在一起就是一个不可估量的大数，无数的小生意集合在一起就是一个不可限量的大市场。

**1.4多样性**

用户的兴趣是多样的，推荐系统需要能覆盖用户各种方面的喜好。这里有个假设，如果推荐列表比较多样，覆盖了用户各种各样的兴趣，那么真实命中用户的兴趣概率也会越大，那么就会增加用户找到自己感兴趣的物品的概率。

**1.5 新颖性**

新颖的推荐是指给用户推荐那些他们以前没有听说过的物品。要准确地统计新颖性需要做用户调查。

**1.6 惊喜度**

如果推荐结果和用户的历史兴趣不相似，但却让用户觉得满意，那么就可以说推荐结果的惊喜度很高，而推荐的新颖性仅仅取决于用户是否听说过这个推荐结果。

**1.7 信任度**

用户对推荐系统的信任程度。如果用户信任推荐系统，那就会增加用户和推荐系统的交互。特别是在电子商务推荐系统中，让用户对推荐结果产生信任是非常重要的。同样的推荐结果，以让用户信任的方式推荐给用户就更能让用户产生购买欲，而以类似广告形式的方法推荐给用户就可能很难让用户产生购买的意愿。提高推荐系统的信任度主要有两种方法。首先需要增加推荐系统的透明度（transparency），而增加推荐系统透明度的主要办法是提供推荐解释。其次是考虑用户的社交网络信息，利用用户的好友信息给用户做推荐，并且用好友进行推荐解释。

**1.8 实时性**

在很多网站中，因为物品（新闻、微博等）具有很强的时效性，所以需要在物品还具有时效性时就将它们推荐给用户。因此，在这些网站中，推荐系统的实时性就显得至关重要。

推荐系统的实时性包括两个方面。首先，推荐系统需要实时地更新推荐列表来满足用户新的行为变化。实时性的第二个方面是推荐系统需要能够将新加入系统的物品推荐给用户。这主要考验了推荐系统处理物品冷启动的能力。

**1.9 健壮性**

衡量了一个推荐系统抗击作弊的能力。算法健壮性的评测主要利用模拟攻击。首先，给定一个数据集和一个算法，可以用这个算法给这个数据集中的用户生成推荐列表。然后，用常用的攻击方法向数据集中注入噪声数据，然后利用算法在注入噪声后的数据集上再次给用户生成推荐列表。最后，通过比较攻击前后推荐列表的相似度评测算法的健壮性。如果攻击后的推荐列表相对于攻击前没有发生大的变化，就说明算法比较健壮。

**1.10 商业目标**

很多时候，评测推荐系统更加注重商业目标是否达成，而商业目标和盈利模式是息息相关的。一般来说，最本质的商业目标就是平均一个用户给公司带来的盈利。不过这种指标不是很难计算，只是计算一次需要比较大的代价。因此，很多公司会根据自己的盈利模式设计不同的商业目标。

**2 新增指标**

我们可以考虑增加/修改 的几个指标：

**1.1 效率性**

所谓效率性，就是 能够以最短时间、最小成本 满足 使用者使用目的的 指标。 我们需要对于 用户的使用目的有一个提前化的度量。

**1.2 平衡性**

在几个 重要指标之间，一个推荐系统是如何达到均衡的，可以提出一个指标判断这个系统是否失衡。