

《推荐系统实践》学习笔记系列(一)

前言

推荐系统解决的问题：信息过载（information overload）。

解决方法：分类-->搜索引擎-->推荐系统

学术界把推荐问题看成评分预测问题，最常见的是TopN推荐。

第一章

1.1什么是推荐系统

- 推荐算法分类

- （1）按数据分

- 协同过滤（social recommendation）

- 内容过滤（content-based recommendation）

- 社会化过滤（social recommendation）

- （2）按算法分

- 基于邻域的算法

- 基于图的算法

- 基于矩阵分解或概率模型的算法

- 什么是推荐系统

- 推荐系统的任务就是联系**用户**和**信息（物品）**，一方面帮助用户发现对自己有价值的信息，另一方面让信息能够展现在对它感兴趣的**用户**面前，从而实现信息消费者和信息生产者的双赢。

1.2个性化推荐系统的应用

(1) 电子商务

亚马逊 (Amazon) 是业界做的最好的之一，其推荐方法主要为协同过滤、内容过滤、打包销售 (cross selling) 。

(2) 电影和视频网站

Netflix是业界做的最好的之一，其推荐方法主要为协同过滤。

(3) 个性化音乐网络平台

个性化音乐电台非常符合个性化推荐技术的产品。

Pandora根据专家标注的基因计算歌曲的相似度，并给用户推荐和他之前喜欢的音乐在基因上相似的其他音乐。

Last.fm主要通过利用用户行为计算歌曲的相似度。

(4) 社交网络

推荐技术主要分三种：

利用用户的社交网络信息对用户进行个性化物品推荐

信息流的会话推荐

给用户推荐好友

Facebook开发了EdgeRank算法对会话进行排序，是用户能够尽量看到熟悉的好友的最新会话。

(5) 个性化阅读

Google Reader让用户关注感兴趣的人，推荐这些人发表的文章。

Zite通过收集用户对文章的偏好信息更新用户的个性化文章列表。

(6) 基于位置的服务

位置是一种很重要的上下文信息，基于位置给用户推荐离他近且他感兴趣的服务，用户更有可能去消费。

(7) 个性化邮件

通过分析用户阅读邮件的历史行为和习惯对新邮件进行重新**排序 (Tapestry) 或分类**

(Google)，从而提高用户的工作效率。

(8) 个性化广告

推荐技术主要分3种：

上下文广告

搜索广告

个性化广告

1.3推荐系统评测

- 推荐系统参与者

	离线实验	问卷调查	在线实验
用户满意度	×	✓	○
预测准确度	✓	✓	×
覆盖率	✓	✓	✓
多样性	○	✓	○
新颖性	○	✓	○
惊喜度	×	✓	×

- 推荐系统评测方法

- (1) 离线实验

优点：不需要真实用户参与，快速，大量。

缺点：无法获取商业上的指标。

- (2) 用户调查

优点：获取用户主观感受

缺点：成本高

- (3) 在线实验

含义：AB测试，用户分组用不同算法进行推荐对比效果

缺点：周期长

- 评测指标

- (1) 用户满意度

- (2) 预测准确度

i. 评分预测

均方误差 (RMSE)

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

平均绝对误差 (MAE)

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

RMSE加大了对预测不准的用户物品评分的惩罚（平方项的惩罚），因而对系统的评测更加苛刻。

ii. TopN推荐

TopN推荐的预测准确率一般通过准确率 (precision) /召回率 (recall) 度量。

令 $R(u)$ 是根据用户在训练集上的行为给用户作出的推荐列表，而 $T(u)$ 是用户在测试集上的行为列表。那么，推荐结果的召回率定义为：

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

推荐结果的准确率定义为：

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

iii.覆盖率

假设系统的用户集合为 U ，推荐系统给每个用户推荐一个长度为 N 的物品列表 $R(u)$ 。那么推荐系统的覆盖率可以通过下面的公式计算：

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

再细化一下,可以用信息熵

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

$p(i)$ 是物品 i 的流行度除以所有物品流行度之和 或者
基尼系数 (Gini Index)

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

这里， ij 是按照物品流行度 $p()$ 从小到大排序的物品列表中第 j 个物品。

iv.多样性

简而言之，就是“不在一棵树上吊死”。

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。假设 $s(i, j) \in [0, 1]$ 定义了物品 i 和 j 之间的相似度，那么用户 u 的推荐列表 $R(u)$ 的多样性定义如下：

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

而推荐系统的整体多样性可以定义为所有用户意见列表多样性的平均值：

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

v.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）

vi.惊喜度

惊喜度与新颖性的区别：假设一名用户喜欢周星驰的电影，然后我们给他推荐了一部叫做《临歧》的电影（该电影是1983年由刘德华、周星驰、梁朝伟合作演出的，很少有人知道这部有周星驰出演的电影），而该用户不知道这部电影，那么可以说这个推荐具有新颖性。但是，这个推荐并没有惊喜度，因为该用户一旦了解了这个电影的演员，就不会觉得特别奇怪。但如果我们给用户推荐张艺谋导演的《红高粱》，假设这名用户没有看过这部电影，那么他看完这部电影后可能会觉得很奇怪，因为这部电影和他的兴趣一点关系也没有，但如果用户看完电影后觉得这部电影很不错，那么就可以说这个推荐是让用户惊喜的。

vii.信任度

度量推荐系统的信任度只能通过问卷调查的方式，询问用户是否信任推荐系统的推荐结果。

提高信任度的主要方法

增加推荐系统的透明度，而增加推荐系统透明度的主要方法是提供推荐解释。

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。

viii.实时性

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。

推荐系统需要能够将新加入系统的物品推荐给用户。

ix.健壮性

概述：任何一个能带来利益的算法系统都会被人攻击，这方面最典型的例子就是搜索引擎。搜索引擎的作弊和反作弊斗争异常激烈，这是因为如果能让自己的商品成为热门搜索词的第一个搜索果，会带来极大的商业利益。推荐系统目前也遇到了同样的作弊问题，而健壮性（robustness）指标衡量了一个推荐系统抗击作弊的能力。

提高健壮性的三种方法

选择健壮性高的算法

设计推荐系统时尽量使用代价比较高的用户行为。比如，如果有用户购买行为和用户

浏览行为，那么主要应该使用用户购买行为，因为购买需要付费，所以攻击购买行为的代价远远大于攻击浏览行为。

在使用数据前，进行攻击检测，从而对数据进行清理。

x.商业目标

总结

	离线实验	问卷调查	在线实验
用户满意度	×	✓	○
预测准确度	✓	✓	×
覆盖率	✓	✓	✓
多样性	○	✓	○
新颖性	○	✓	○
惊喜度	×	✓	×

*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：包括物品的属性信息、流行度、平均分以及是不是新加入的物品等。

实践维度：包括季节，是工作日还是周末，是白天还是晚上等。

本着能把书读薄的意愿写的技术报告。希望以后能越写越好。如有建议可以通过
qhd0081@vip.163.com联系我。

作者[钱昊达]

2018年8月14日