

## 复盘 2 | 推荐系统核心技术模块

2018-04-15 洪亮劫

AI技术内参

[进入课程 >](#)



到目前为止，我们讲完了人工智能核心技术的第二个模块——**推荐系统**。

整个模块共**21期**，**7大主题**，希望通过这些内容，能让你对推荐系统核心技术有一个全面系统的认识和理解，为自己进一步学习和提升打下基础。今天我准备了 21 张知识卡，和你一起来对这一模块的内容做一个复盘。

提示：点击知识卡跳转到你最想看的那篇文章，温故而知新。如不能正常跳转，请先将 App 更新到最新版本。

### 现代推荐架构剖析

推荐架构需要解决的问题：

能够在一两百毫秒内给用户提供当前的推荐结果；

对用户和系统的交互结果做出响应；



## 基于线下离线计算的推荐架构

线下离线计算的一个主要想法就是：把计算中复杂的步骤尽量提前做好，然后当用户来到网站需要呈现结果的时候，我们要么已经完成了所有的计算，要么还剩非常少的步骤可以在很快的时间内，也就是所说的一两百毫秒内完成剩下的计算。

## 基于多层搜索架构的推荐系统

多层搜索架构可以支持搜索结果，自然地，对实时呈现推荐结果有很好的支持。

需要对用户的反馈进行更新，可以在重排序的阶段，通过两种方法实现：其一是更新重排序阶段的模型，其二是更新重排序的模型的某些特性。

搜索架构对于新用户是天然支持的，但对新物品的支持是其短板。

## 复杂现代推荐架构漫谈

当我们面对新用户多和新产品多的场景时，推荐架构的一些基本规则：

- 尽可能把复杂的运算放在线下，因为毕竟需要在规定的时间内返回结果；
- 在一切有可能的情况下，尽可能使用搜索引擎来减少需要对大量物品进行打分的步骤；
- 对于活跃的用户，我们可以使用多层搜索架构；但是对于不活跃用户，我们可以依赖线下，提前产生所有的推荐结果。

### 简单推荐模型

## 基于流行度的推荐模型

物品流行度影响因素：时间和位置。

对于流行度的衡量，我们往往使用的是一个“比值”，或者是计算某种“可能性”。

建立无偏差的数据：epsilon贪心、最大似然估计法。



## 基于相似信息的推荐模型

相似信息的推荐模型又叫“临近”模型，其内在假设是“协同过滤”，就是相似的用户可能会有相似的喜好，相似的物品可能会被相似的人所偏好。

“协同过滤”从统计模型的意义上来讲，其实就是“借用数据”，在数据稀缺的情况下帮助建模。

## 基于内容信息的推荐模型

基于内容信息的推荐系统，其实就是用特征来表示用户、物品以及用户和物品的交互，从而能够把推荐问题转换成为监督学习任务，有两个关键步骤：特征工程、目标函数。

### 内容信息的各类特性：

- 物品的文本信息；
- 物品的类别信息（或者物品的知识信息）；
- 用户的基本特性，包括性别、年龄、地理位置；
- 用户画像。

## 基于隐变量的模型

我们通过模型的假设，知道隐变量之间的关系，但暂时并不知道隐变量的取值。因此需要通过“推断”过程来确定隐变量的实际取值。当我们知道了这些隐变量的取值之后，就可以根据这些取值来对未来的数据进行预测和分析。

隐变量往往还带有“统计分布”的假设。最简单的隐变量模型是高斯混合模型。

## 矩阵分解

---

在推荐系统中，我们可以得到一个矩阵。这个矩阵的每一行代表一个用户，每一列代表一个物品，每一个交叉的元素代表某一个用户对于某一个商品的评分。这个矩阵最大的特点就是数据非常稀少，需要补全整个矩阵里的其他元素。

学习矩阵分解模型，一种常用的方法就是利用最小二乘法的原理来拟合求解。

## 基于回归的矩阵分解

矩阵分解的基本模型存在两个问题：其一矩阵分解的矩阵仅仅是对用户和物品的喜好进行了“编码”；其二矩阵分解的核心是学习用户的隐向量和物品的隐向量。

基于回归的矩阵分解假定，用户的隐向量，其实是从用户的显式特性变换而来的。同理，物品的隐向量，其实是从物品的显式特性变换而来的。而这两种变换，本质上就是两个回归模型。



## 分解机

分解机结合了“基于内容的推荐系统”和“基于回归的隐变量模型”的一些基本思想。

在实际应用中，我们经常使用“随机梯度下降”来对分解机直接进行求解。





## 张量分解模型

---

从本质上来说，张量就是矩阵的推广。矩阵是对二维关系建模的一种工具；而张量，就是对N维关系的一种建模。

张量分解至少有两种不同的形式，分别是“CP分解”和“HOSVD分解”。

两种求解张量分解的方法：“随机梯度下降”法和ALS方法。

## 协同矩阵分解

协同矩阵分解的基本思路，就是有多少种二元关系，就用多少个矩阵分解去建模这些关系。

在协同矩阵分解的场景中，学习各个隐变量的参数的过程，和一般的单个矩阵分解相比，没有太多本质性的变化。最简单的学习过程，依然是利用“随机梯度下降”法去学习。



## 优化复杂目标函数

把搜索系统中的排序思想利用到推荐系统中是推荐系统的一个重大进步，这也让推荐系统和真实场景逐渐挂钩。

要想更改推荐系统的行为，从评分的预测到排序学习，我们需要更改目标函数。



## EE算法综述

---

EE可以看作是一个优化过程，需要多次迭代才能找到比较好的方案。

EE的产品部署两大难点：如何上线测试、如何平衡产品。



## UCB算法

---

UCB算法本身，其实是同时考虑了物品现在的情况以及在这种情况下的置信度，并且寄希望通过多次迭代来达到减小标准差，提高置信度的目的。

UCB算法本质上还是“确定性”算法，并没有随机性。

## 汤普森采样算法

### 汤普森采样算法核心要点：

- 每一轮，汤普森采样都有一个参数采样的动作；  
从后验概率分布中进行抽样；
- 因为使用了贝叶斯统计，对参数有先验的设置，  
因此针对当前点击率估计还不准确甚至还没有数据的物品来说，有天然的优势；
- 因为是采样，即便是在参数一样的情况下，两个物品的采样数值都有很大可能是不一样的，一举解决了UCB的问题。

### 基于深度学习的推荐模型

## 受限波兹曼机 (RBM)

RBM就是由一层隐单元和一层显单元组成的神经网络结构。

RBM对于推荐系统的建模看上去很简单，但是难点却是如何学习这些未知的权重。



## 基于RNN的推荐系统

RNN的输入是当前的一个物品，然后RNN需要输出的是对于下一个物品的预测，同时为了能够对这些在时间序列上的物品进行建模，RNN内部也要维持一个隐含的状态序列，这个状态序列会随着时间的变化而变化。

## 利用深度学习来扩展推荐系统

多层神经网络是一种提取特性的利器。直接利用多层神经网络对用户和物品的建模可以简单归纳为两步：首先，把离散的ID信息转换成为连续的信息，形成嵌入层；然后，利用多层神经网络对嵌入层进行变换，并最后输出预测结果。

其他深度学习模型在推荐系统中的应用尝试，比如自动编码器和CNN。

### 推荐系统的评价



## 传统线下评测

基于评分的线下评测：均方差、方差。

基于排序的线下评测：精度、召回。

利用排序的思路来评测推荐系统，已经成为了目前推荐系统线下评测的一个标准指标。

## 线上评测

如何能够有效地进行在线实验，包括实验设计、实验的评测等，都是非常前沿的研究课题。

### 通用的推荐系统线上评测指标：

- 用户的驻留或者停留时间；
- 用户在相邻两次访问中的间隔时间，有时叫作“空缺时间”。



## 无偏差估计

在有偏差的系统中，先通过数据学习得到模型，然后再部署到系统中去，这个流程其实严重阻碍了我们对用户真实喜好的检测。因此，这也是线下表现和线上表现不一致的一个原因。

进行无偏差估计的一个基本假设和要求：需要假设收集的数据涵盖了整个数据集。

最后，恭喜你在这个关卡下已经阅读了 45397 字，听了 138 分钟的视频，获得一张新的**通关卡**，这是一个不小的成就。在人工智能领域的千里之行，我们又往前迈出了一步。

## 恭喜你！获得一张“内参”通关卡

### 🕒 你已学习

21 期 | 45397 字 | 138 分钟

### 🚩 8个关卡

- |                 |   |
|-----------------|---|
| 第一关：搜索          | ★ |
| 第二关：推荐系统        | ★ |
| 第三关：广告系统        | ★ |
| 第四关：自然语言处理及文本处理 | ★ |
| 第五关：计算机视觉       | ★ |
| 第六关：人工智能国际顶级会议  | ★ |
| 第七关：数据科学家养成     | ★ |
| 第八关：数据科学团队养成    | ★ |

AI技术内参

感谢你在专栏里的每一个留言，给了我很多思考和启发。期待能够听到你更多的声音，我们一起交流讨论。



# AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管  
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 083 | 基于深度学习的推荐模型之三：利用深度学习来扩展推荐系统

下一篇 084 | LDA变种模型知多少

## 精选留言 (2)

 写留言



永夜

2018-04-18

感谢老师

展开 ∨

 4



Junjian

2018-11-17

谢谢，获益良多

展开 ∨





下载APP

