

# R&S | 手把手搞推荐[0]: 我的推荐入门小结

原创 机智的叉烧 CS的陋室 2019-05-07



点击上方蓝色文字立刻订阅精彩

## The Next Episode

San Holo;Snoop Dogg;Dr. Dre - The Next Episode (San Holo Remix)



### 【R&S】

本栏目从原来的【RS】改为【R&S】，意为“recommend and search”，即推荐和搜索，结合本人近期的工作方向、最近上的七月在线的课、自己自学、而退出的特别栏目。当然，按照我往期的风格，更加倾向于去讨论一些网上其实讲得不够的东西，非常推荐大家能多看看并且讨论，欢迎大家给出宝贵意见，觉得不错请点击推文最后的好看，感谢各位的支持。

另外，【手把手搞推荐】是我近期开始的连载，结合自己所学，带上代码的手把手和大家分享一些模型和数据处理方式，欢迎关注。

往期回顾：

- [RS | 推荐系统整体设计](#)
- [R&S | 爱奇艺搜索启发](#)
- [RS | 深度讨论FM和FFM：不仅是推荐](#)
- [RS | 论文阅读：用于YouTube推荐的深度神经网络](#)
- [NLP.TM | 我的NLP学习之路](#)

去年年末至今，学习推荐系统已经接近半年，在各种事情的忙碌下，依然坚持着完成一些自学，推荐系统的入门流程逐渐到达尾声，自己学的内容也逐步完善，门是基本入了，所以打算写一个小系列，在自己的设备、数据条件允许的前提下，把一些关键但是基本的模型给实现一遍。难得、前沿的可能还有待进一步学习，有问题欢迎指正。

## 学习历程

至于为什么想学推荐系统，可以看看我这篇历史文章：技术向：推荐学习推荐系统（深度思考，不是广告）。

最开始的学习是想看项亮老师的《推荐系统实践》，相信很多前辈也从这本书里受益匪浅，这本书我是基本看完了，但是感觉意犹未尽，在深度学习如此流行的环境下，推荐系统肯定不止于此，于是砸钱报了一波七月在线的推荐系统课。

七月在线的课吧，其实我一直对这种基本的课没有太高的要求，不希望从课堂上真正学到什么，而是让我知道，我该学什么，这门课肯定是做到的，虽然后续的课程因为很多原因我没有跟上，但是在录屏和PPT的帮助下，我把老师课上说的，以及老师提到的资料都认真看了，我想这才是我最大的收获。

目前的状态是，理论基本都能get，架构也基本理解，甚至有一定能自己弄一个简单的架构的思想萌芽，对整个流程有比较深入的理解，相信这是一个比较重要的里程碑吧，缺的可能是动手吧，在未来很短的一段时间内，我会集中一些精力在一套数据下尝试一些基本的模型。

## 入门策略建议

我想很多人肯定是想知道我的学习建议，我就来谈一下吧。

首先，恳请各位想入门机器学习的同学，认认真真先把机器学习搞懂，推荐里面大量机器学习、深度学习模型。还有很多有关数据层面、特征层面的工作，所以不要指望算法小白能够直接看懂里面的东西，这点在我看来推荐系统并不友好。

第二，从协同过滤入手，看一些基本的理论，在召回层面了解一些基本的知识，适当情况可以动手实践一下。

第三，排序阶段，从CTR预估之类的典型问题的处理，如LR、FM、FFM系列模型，到wide&deep等，另外是矩阵分解，对ID类特征的处理，借用NLP中embedding的一套深度了解，甚至到一些更前沿的模型，铺开了解并且掌握。

第四，在这些都基本了解的基础上，去看一些企业的经验和case，如我之前写过的Youtube、爱奇艺等，深入了解，看看企业级的是怎么做的，不过放心，他们告诉你的一定是他们基本已经不用的东西，顶多是第一版上线的基线模型，尽管如此，对于新手、从无到有建模而言，已经绰绰有余。

第五，非常建议大家看看论文，尤其是顶会的、著名企业的论文，多去看看学学，这些都是面向实践非常重要的论文；对于学术界，就针对自己的问题去找论文吧，记得最好从顶会下手。

我的学术研究方向不在这块，在NLP，因此推荐系统的学术界了解不如各位大佬那么深，就不继续深说了。

补充一句，抛开实现谈理论，在这块肯定不可行的，纸上得来终觉浅，绝知此事要躬行，动手敲几行代码吧。

## 学习材料

很多人问我有什么好的学习材料其实我自己非常头疼，因为我不知道对你来说什么是好的学习材料，也不知道你的基础怎么样，直接给你弗朗西斯科的你可能要捶我对吧，虽说现在推荐系统还在开荒阶段，学习资源少，但是从百度等完了过去到终究是找得到资源的。

无差别的，我说几个比较建议大家看的材料吧。

## 书籍

- 《推荐系统实践》，项亮。以协同过滤为基点去写的书，对实际业务也有很多独到的理解，冷启动等，说的很清楚，适合对背景不太熟悉的同学。
- 《推荐系统-技术、评估及高效算法》，弗朗西斯科等。推荐系统界的算法导论，我就说到这吧，见仁见智。
- 《推荐系统与深度学习》，清华大学出版社那本。在书里面，讲的可以说非常前沿了，对一些具体问题怎么处理讲的十分详细，还有代码，基线模型基本都有了，可以参考。
- 《计算广告》，刘鹏等。这本书业内也非常出名，对广告领域、推荐领域的很多问题都说的很详细，甚至是和产品等有关工作者看。
- 《用户网络行为画像》，牛文佳等。简单提到推荐系统，但是在推荐系统这种面向用户、复杂特征的问题，还是非常适用的。

多说无益，重要的是看完吧。

## 其他材料

说到课，别的没上过，七月在线的推荐系统课还算合适，说不定你们还能在群里看到我哈哈哈（别加我QQ，基本不用），不过提醒，指望一门课几十个小时就把你讲懂了都会了，做梦，自己去认真看老师给的材料，自己动手做，才会有搜获，别整天等着别人把东西往你的嘴里喂。这门课，基本知识在我的视角下都讲到了。（没收广告费）

论文吧，经典论文多看看，上面提到了此处不赘述。

## 我打算写什么

这个系列归属于我【R&S】专栏里面，当然是写怎么搞推荐系统，那我具体想些什么呢。

- 既然手把手，肯定是写一些实操性的东西了。
- 数据探索与基本的数据分析，这是很容易忽略但是却非常重要的工作，也有人问我数据探索怎么做，我就简单操作一次吧。

- 一些偏向机器学习、深度学习模型吧，LR、FM等的尝试，一些关键的模型吧，都尝试尝试，看看综合性能有什么差别。
- 一些特征工程方法，看看性能提升量，以及具体操作细节，one-hot化、embedding等。
- 有些可能对数据要求比较高的，可能干不了，设备和数据本身不允许，可以尝试吧。

到这里吧，本系列下一篇是有关数据探索和数据介绍的内容，后续系列主要是用movielens中的ml-1m的数据，有兴趣的可以去了解一下。

## 我是叉烧，欢迎关注我！

叉烧，机器学习算法实习生，北京科技大学数理学院统计学研二硕士（保研），本科北京科技大学信息与计算科学、金融工程双学位毕业，硕士期间发表论文5篇，学生一作3篇，1项国家自然科学基金面上项目学生第2参与人，参与国家级及以上学术会议4次，其中，1次优秀论文，国家奖学金。曾任去哪儿网大住宿事业部产品数据，美团点评出行事业部算法工程师。



微信个人公众号  
CS的陋室

微信                      zgr950123  
邮箱                      chashaozgr@163.com

喜欢此内容的人还喜欢

属于算法的大数据工具-pyspark：10天吃掉那只pyspark

CS的陋室

反思中国互联网特色996：不应要求每一个员工都变成马云

InfoQ

初一“铁”、初二“钢”、初三才能“响当当”！附：初中三年提升攻略

简单初中生