



下载APP



开篇词 | 从0开始搭建一个深度学习推荐系统

2020-09-21 王喆

深度学习推荐系统实战

[进入课程 >](#)**讲述：王喆**

时长 13:21 大小 12.24M



你好，我是王喆，在一家硅谷的科技公司 Roku 担任机器学习工程师。

Roku 是美国最大的视频流媒体平台，在美国智能电视市场占比超过 1/3。我作为推荐系统的架构负责人，在最近两年多的时间里，跟团队一起搭建了一整套深度学习推荐系统，相比于传统推荐系统，它在播放时长、点击率等效果指标上取得了 30% 以上的提升。

在这个过程中，我和团队踩过不少“坑”，也总结了很多经验。其中最宝贵的一点是，**只有建立起深度学习推荐系统的知识体系，从系统的层面考虑问题，我们才能够实现整体效果上的优化。**与此同时，我还发现越来越多的在校生的和刚入行的工程师，想要熟悉或是在深度学习推荐系统领域取得进一步的发展，但经常因为缺少系统的学习机会而困在一两个难点上，停滞不前。



比如说，一些对**深度学习推荐系统非常感兴趣的在校生**，很希望能有前辈带领他们熟悉业界的热门技术，让他们在进入相关领域前就积攒起足够的工程经验。这种想法是非常好的。我也面试超过两三百位应届生了，以我的经验，最让企业青睐的应届生就是动手能力强、有实习经验、有业界项目经验的。但这种锻炼自己的机会却很难找到。

再比如说，一些已经进入**推荐、广告、搜索相关部门，但是时间不长，或者正要转行进入这些行业的工程师**。他们想要自主承担起某个模块的开发任务，期待扩展自己的技术视野和格局，向更高级的职位进发。但这需要他们建立一套成体系的知识和实践框架。

我是一个十分赞同开源思想并且乐于分享知识的人。既然有那么多同行都有这样的需求，我为什么不把我建立行业知识体系的过程分享出来呢？所以在 2020 年初，我发布了新书《深度学习推荐系统》，希望这本书能帮助一些同学建立深度学习推荐系统的知识体系。令我没想到的是，这本书在半年之内就重印了 7 次，销量接近 2 万册，豆瓣评分达到了 9.3，这对于一本技术书来说是非常难得的。

新书获得肯定的同时，我也收到了很多读者的反馈，最多的一点是“书中的知识非常体系化，帮助我建立了整个行业的知识蓝图，但如果能有配套的实践项目就更完美了”。

就像 Linux 之父 Linus Torvalds 说的那句话：“**Talk is cheap. Show me the code.**”，实践对于工程师来说永远是最重要的。所以我就趁热打铁，在极客时间上开设了一门更偏重实战的技术专栏。我想这是一次绝佳的机会，**与你一同从“0”开始，搭建一个“工业级”的“深度学习”推荐系统**，做到知识和实践两手抓！

深度学习的浪潮，推荐系统的时代

在跟你聊了开设这门课程的初心之后，我觉得很有必要把我们的视野放得更宽一点，看看我们这个时代到底是怎么被推荐系统影响着，推荐系统又是怎么在深度学习的浪潮之中“乘风破浪”的。

毫无疑问，推荐系统从来没有像现在这样影响着我们的生活。想上网购物，天猫、京东的推荐系统会帮你挑选商品；想了解资讯，头条、知乎的推荐系统会为你准备感兴趣的新闻和知识；想消遣放松，抖音、快手的推荐系统会为你奉上让你欲罢不能的短视频。

而驱动这些巨头进行推荐服务的，都是基于深度学习的推荐模型。

2013 年，百度率先在广告系统中应用了深度学习，2015 到 2020 年，阿里提出并应用了从 MLR 到 DIEN 等一系列的深度学习模型。国外的互联网巨头也不遑多让，从最早的 Google 的 Word2vec，到 2015 年 YouTube 的深度学习推荐系统，再到之后的 Facebook、Amazon、微软等等，几乎所有头部公司的成功应用，让深度学习如后浪般席卷了推荐系统业界，将传统的推荐模型彻底取代。

更让人感叹的是，字节跳动在 2020 年 1 月 5 日发布的一份《抖音数据报告》中宣布，抖音日活用户突破了 4 亿。要知道，这距离抖音 2016 年 9 月上线仅过去了 3 年多一点的时间。作为一个几乎完全由推荐系统驱动的应用，这样的增长速度是惊人的，是前所未有的。而字节跳动技术团队曾经披露的，深度学习在推荐算法、视频内容理解、自然语言处理等方向上的应用，则又一次向我们印证了深度学习的强大实力。

深度学习时代，推荐工程师的职业发展方向在哪？

随着推荐系统的快速发展，在深度学习时代，推荐工程师又该如何选择自己的职业发展方向呢？

推荐工程师的工作，本质上是利用一切可能的技术手段来提升推荐系统的效果，从而不断达到甚至超越企业的商业目标。

举个例子，2019 年，阿里著名的千人千面系统驱动了天猫“双 11” 2684 亿元的成交额。假设我们通过改进天猫的商品推荐功能，让平台整体的转化率提升 1%，那么在 2684 亿元成交额的基础上，我们就能再增加 26.84 亿元。也就是说，推荐工程师仅通过优化推荐技术，就创造了 26.84 亿元的价值。这无疑是这个职位最大的魅力所在，也是它能够支撑起百万年薪最重要的原因。

1% 听起来很小，但是想要在一个成熟的推荐系统上，找到能够提升的突破点并不容易，这需要我们动用所有的领域知识储备。一位推荐工程师优秀与否，也就是在这个时候体现出来的。

在所有业界巨头的推荐引擎都由深度学习驱动的今天，作为一名推荐系统从业者，我们不应该止步于，或者说满足于继续使用协同过滤、矩阵分解这类传统方法，而应该**加深对深度学习模型的理解，加强对大数据平台的熟悉程度，培养结合业务和模型的技术直觉，提高我们整体的技术格局，这些都是我们取得成功的关键。**

可能你最近经常听说“算法工程师的知识更新太快，一不小心就处在被淘汰的边缘”。但我始终坚信一点，人才的分布都是金字塔式的，与其抱怨金字塔底座为什么这么宽，不如努力提高自己的知识储备、工程能力和技术视野，哪怕只是向金字塔的塔尖前进了一步，我们也超越了最宽的那个底座，不是吗。

因此，我希望在这门课里与你一起建立的，是深度学习推荐系统的整体架构。这对我来说也是个不小的挑战，因为它几乎需要我输出所有的工作经验和知识储备。当然，对你来说肯定也是个挑战，因为一定有很多新知识需要补充到自己的知识框架中。但我相信我们的目标一定是一致的，就是在深度学习时代，在推荐系统这个行业开拓视野、站稳脚跟。

这门课是怎么设计的？

一句话来说，这门课程是一门知识与实践并重的课程，通过解决 30+ 个深度学习推荐系统问题，不仅能串联起深度学习推荐系统的知识体系，还能帮你实打实地收获一套经过实践验证过的开源代码，从而让你也能实现一个工业级的深度学习推荐系统。

注意了，这里面有几个关键词，分别是“知识体系”“深度学习”“工业级”“实战”。没错，我们的课程就是围绕这几个关键词展开的。我遵循一个经典推荐系统的框架，把课程分为 6 个部分，分别是“基础架构篇”“特征工程篇”“线上服务篇”“推荐模型篇”“效果评估篇”“前沿拓展篇”，其中的每节课，我们都会着重解决一个技术难点。

基础架构篇：从 0 出发，建立深度学习推荐系统的知识体系

在开始学习这门课之前，我对你的要求有两个，一是有一定的编程基础，二是有基本的机器学习概念知识。在此基础上，我们通过基础架构篇的学习，就能建立起深度学习推荐系统的完整知识架构，做到“心中有蓝图、心中有高楼”。

具体来说，在基础架构篇中，我会详细讲解我们要从 0 开始实现的推荐系统，Sparrow RecSys 的主要功能和技术架构。由于缺少工业级的实验环境，Sparrow RecSys 不可能是一个真正的工业级推荐系统，但是它的每一行代码都是严谨的，其中的每个方法都是经过业界验证的主流方法。并且，我们还会使用到 Spark、Flink、TensorFlow 这些业界目前最流行的机器学习和大数据框架，麻雀虽小，但五脏俱全。

特征工程篇：又快又好，用心准备推荐系统的“食材”

在特征工程篇中，我会和你一起讨论推荐系统会用到的特征，以及主要的特征处理方式，并且把它们都实践在 Spark 上。除此之外，我还会讲解深度学习中非常流行的 Embedding、Graph Embedding 技术。

我们可以把特征工程看作是为推荐系统准备“食材”的过程。所以我希望通过这部分的学习，你不仅能够成为一名合格的“备菜”师傅，更能够利用学到的 Embedding 方法，来实现 Sparrow Recsys 中的相似电影推荐功能，在实践中快速成长起来。

线上服务篇：实践出真知，掌握搭建工业级推荐系统的核心技能

一个工业级推荐系统和实验室 Demo 的最大区别就在于线上服务部分。在这一篇中，我们要实打实地搭建一个推荐服务器，它包括了服务器、存储、缓存、模型服务等相关知识。相信通过这部分的学习，你能初步掌握 Jetty Server、Spark、Redis，这些工程领域的核心技能。

推荐模型篇：深度学习推荐系统上的明珠

如果让我挑出深度学习对传统推荐系统最大的改进，毫无疑问是深度学习在推荐模型上的应用，甚至我们称它为“推荐系统上的明珠”也不为过，所以这一部分可以说是整门课程的重中之重了。我们将一起学习深度学习推荐模型的原理和实现方法，主要包括 Embedding+MLP、Wide&Deep、PNN 等深度学习模型的架构和 TensorFlow 实现，以及注意力机制、序列模型、增强学习等相关领域的前沿进展。

效果评估篇：建立成体系的推荐系统评估机制

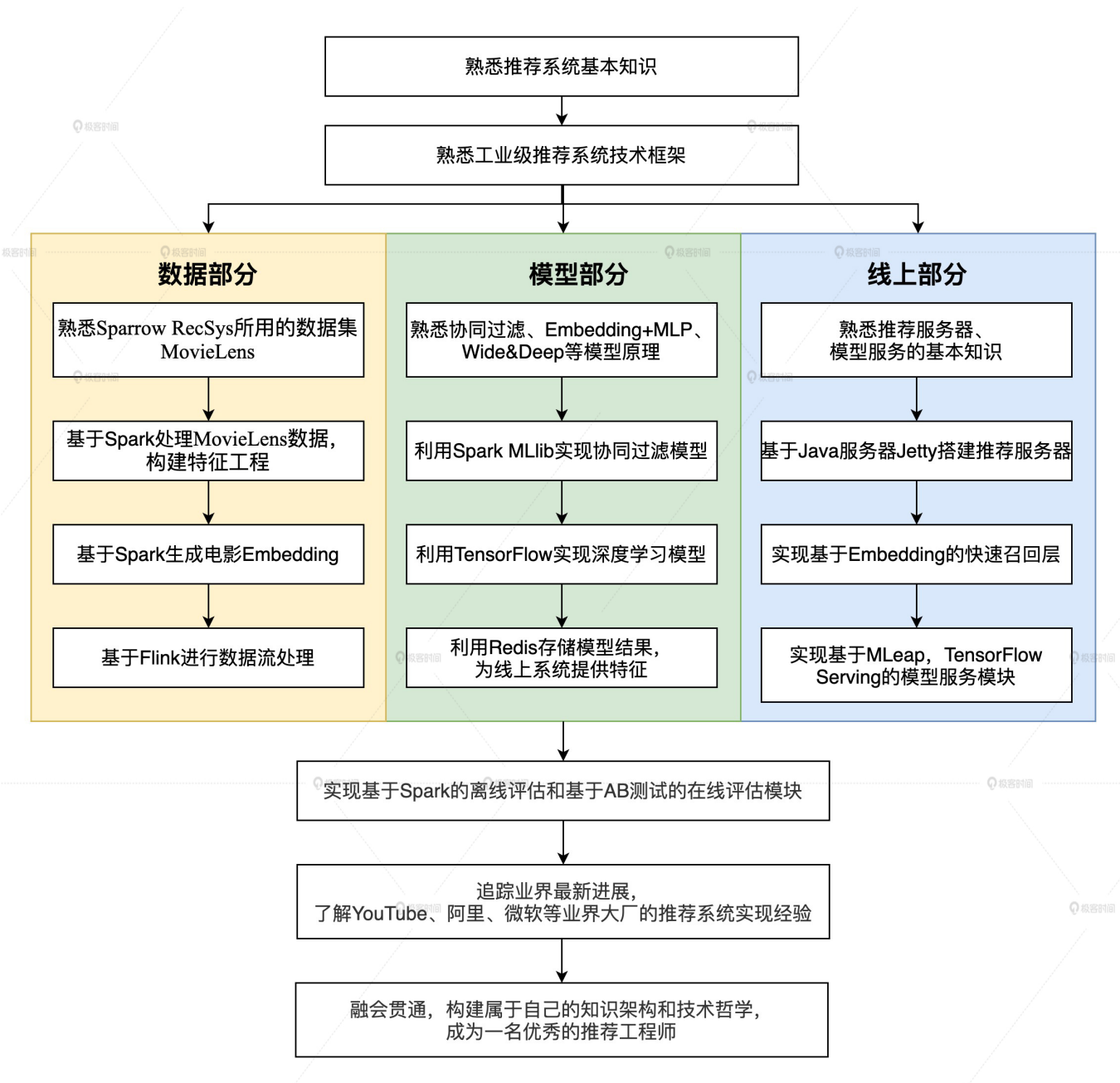
在效果评估篇中，我们要重点学习效果评估的主要方法和指标。但对一个成熟的推荐系统来说，仅熟悉这些是不够的，我还期望你能通过这一篇的学习建立起包括线下评估、线上 AB 测试、评估反馈闭环等整套的评估体系，真正能够用业界的方法而不是实验室的指标来评价一个推荐系统。

前沿拓展篇：融会贯通，追踪业界前沿

在完成整体的知识积累之后，我们在通过这一篇的学习，将通过业界巨头们的深度学习推荐系统方案进行融会贯通。我会重点讲解 YouTube、阿里巴巴、微软、Pinterest 等一线

公司的深度学习应用，帮助你追踪业界发展的最新趋势，并且找到自己技术道路上的方向。

所有的点穿成线、连成面，就组成了我们希望掌握的深度学习推荐系统架构。与此同时，我们也会在课程完成后，搭建起一个完整的推荐系统，再收获一份亲眼看见自己学习成果的成就感。



最后我想说，No Magic，不要期望一门课程就能够让你成为业界专家。如果你是完全没有推荐系统基础的新人，这门课程能够让你入门推荐系统，初步掌握深度学习推荐系统各模块的相关知识和业界实践。如果你是行业老兵，这门课能让你查漏补缺，在技术视野和格局上有所提高。但推荐系统的每个模块都有着极深的技术纵深，不管是 TensorFlow 还是 Spark，还是 Redis、Flink，它们中的每一个都需要我们持续性的钻研才能够成为领域专

家。在技术专家的道路上，这仅仅是个开始，当然，我也很荣幸能帮你开启这个既有魅力又有挑战性的技术领域。

最后，关于深度学习推荐系统，希望你能在这里畅所欲言，提出你的困惑和疑问。也欢迎多多给我留言，你们的鼓励是我的动力。如果你身边也有想要学习深度学习推荐系统的同学，也别忘了把这个课程分享给他。**很高兴能与你一起开启深度学习推荐系统的学习之路，愿与你在攀登行业塔尖的路上共勉。**

50 人觉得很赞 | 提建议

更多学习推荐

机器学习训练营

成为能落地的实干型机器学习工程师

王然 众微科技 AI Lab 负责人

前100名秒杀 ¥3649  加赠书籍



© 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。

下一篇 01 | 深度学习推荐系统的经典技术架构长啥样？

精选留言 (25)

 写留言





huyy6789

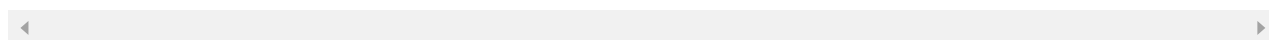
经济学专业，只有python基础和统计学习基础可以看吗？

作者回复: 这门课会涉及到的主要语言是java, scala和python。另外需要一些机器学习基础。所以可能会稍微有些困难。

但我基本会用“从0开始”的语言来讲解这门课程，希望把门槛降下来。

而且在第一篇的最后我会给完全没有这个领域基础的同学一个参考书列表。如果能够基本过一两本参考书，把基础打好，是完全没有问题的。

还是推荐挑战一下自己，祝好。



7



林深时见鹿 🌲 🌲

2020-09-24

老师，除了这门课，有什么推荐的读物吗？

展开 ∨

作者回复: 有的，国庆节前会发一个番外篇，给大家介绍一些推荐的参考书。



3



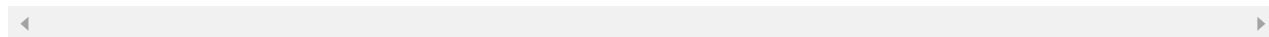
Daryl

2020-09-22

王喆老师的课必须盲买👏

展开 ∨

作者回复: 那就多谢支持啦！



3



野水晶体

2020-09-23

追书而来～

展开 ∨

作者回复: 多谢支持！



2

**georgesuper**

2020-09-22

挺好的课，我是资深大数据工程师，业余学了几年机器学习，老师这课可以让我的知识点贯通升华

作者回复: 基本是我开设这门课的初衷之一，希望能帮助到想建立起全局推荐系统框架的同行们学习提高。



2

**叮当小刚**

2020-09-26

深度学习推荐系统这本书看了百分之八十了，感觉很不错，很有体系，看了以后有一种对整个推荐系统领域的大致轮廓了然于心的感觉，

展开 ∨

作者回复: 多谢支持。希望这门课能让你对推荐系统有更深入的理解!



1

**Geek_66666**

2020-09-23

搭建线上服务器部分，单机有办法搭建吗

展开 ∨

作者回复: 可以，只不过是在单机运行，在SparrowRecSys里可以直接看到效果。与工业级线上服务器的差别在于无法体验多节点负载均衡，这必须依赖工程环境了。



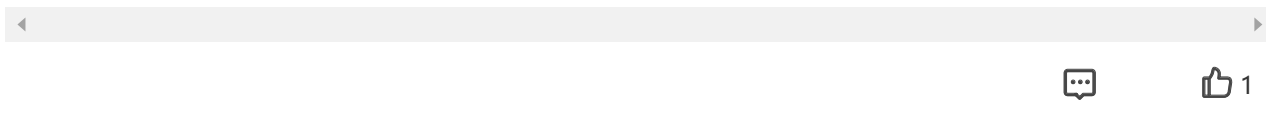
1

**少刷票圈多读书**

2020-09-21

追完了王喆老师的知乎专栏、《百面机器学习》，正在学习《深度学习推荐系统》，就出了实战课程，真香~

作者回复: 非常感谢对这所有作品的支持, 期待提出更多建议!



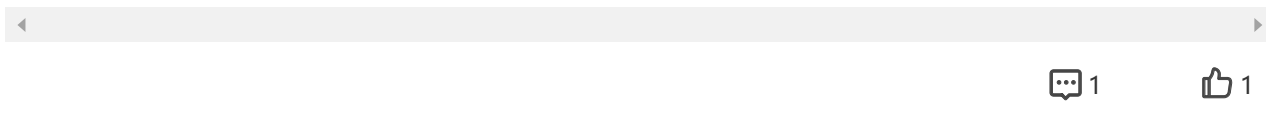
瓜瓜

2020-09-21

不知道 深度学习 在复杂业务场景下 有没有应用场景。

作者回复: 要看一下具体是什么复杂业务场景, 如果是比较复杂的推荐场景的话, 一般采用深度学习主模型, 混合多种推荐策略的方式满足业务需求。

深度推荐模型肯定不是全部, 一定要跟具体的业务逻辑结合在一起才能支持实际的需求。

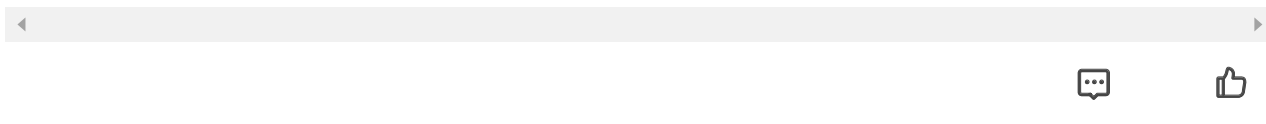


lianlian

2020-10-26

早几周就买了这个课, 今天才有时间听, 接下来会加快速度跟上最新课程。刚好需要, 感谢老师! 突然发现, 您就是我手上这本《深度学习推荐系统》的作者哈哈
展开 ∨

作者回复: 加油



Geek_358a15

2020-10-22

期待

展开 ∨

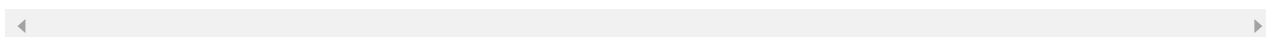


蓝蓝蓝

2020-10-22

现实工业中推荐系统用Java搭建的吗? 还是c++呢 比较基础的问题嘿嘿

作者回复: Go,Java,C++都有。Java对于大家来说比较友好, 所以这门课选择了Java进行实现。



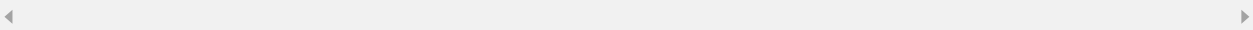
**errocks**

2020-10-16

1. 想请问有没有群，可以实时讨论下问题，交流的
2. 想请问能不能给个资源贴，学习路线

作者回复: 1.目前没有。

2.这就是开设本门课的主要意义吧。

**Wiiki**

2020-10-10

王老师，之前接触过常用的深度学习算法，例如cnn，rnn，word2vector等算法原理，但是没有对一个具体领域的深入实践过。想问一下，是否可以通过学习推荐系统涉及的深度学习相关算法来对深度学习核心概念有个整体掌握呀？

展开 ∨

作者回复: 肯定可以增加一些应用和实践经验，比如各种模型结构的区别和联系，以及为什么产生这种区别和联系

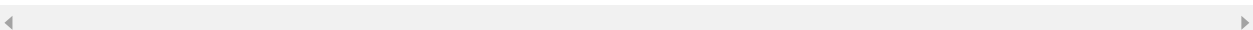
**AtlasGcx**

2020-09-30

在fb做广告，对推荐系统感兴趣所以来学（而且觉得老做广告业障重哈哈），买了老师的书觉得很棒！

作者回复: 推荐本质上也是为了间接提高广告收入，要这么说业障都很重。。

fb有很多优秀的广告和推荐的工作，有机会也期待多分享。

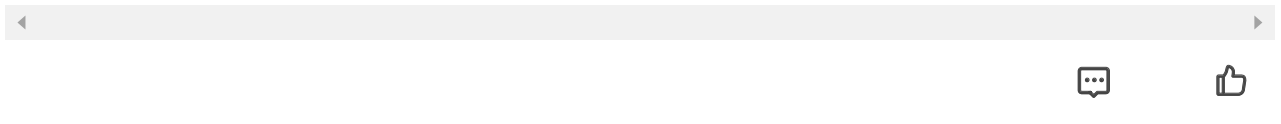
**空白**

2020-09-25

暑期实习就是靠着《深度学习推荐系统》入门的。发现真的喜欢上了这个方向。但是学校实验室缺少这个环境。希望能在这里更好的学习。

展开 ∨

作者回复: 可以把这个项目当作一个种子项目, 项目中使用的movieLens数据集也是学界都公认的权威数据集。

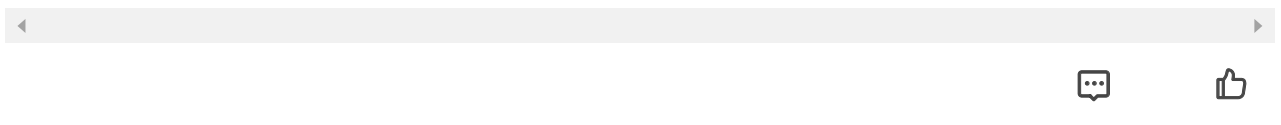


太子长琴

2020-09-25

支持支持, 希望学完自己可以搭建一个出来
展开 ∨

作者回复: 如果完整的学习下来肯定可以!

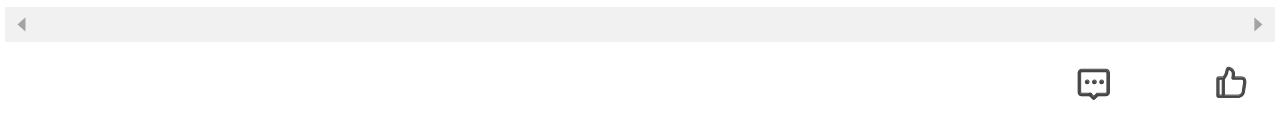


成锋

2020-09-24

有编程基础, 没有机器学习的概念, 学起来怎么样呢
展开 ∨

作者回复: 我觉得跟下来肯定是没有问题的, 因为我基本会用易懂的语言来讲解。但是想深入理解还是要有机机器学习的基础, 我之后会推荐大家一些参考书, 根据自己的知识基础, 可以买一两本来补充一下。

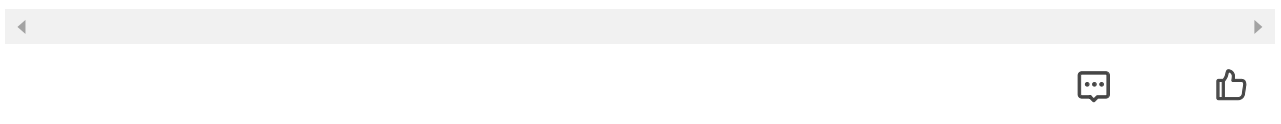


薛定谔的疯兔子

2020-09-24

老师的《百面机器学习》《深度学习推荐系统》我都常常有翻阅。虽然现在从事的还是一般的java后端开发工作, 但是对ml,dl, 还有推荐领域很有兴趣, 希望自己能多和老师学到一些具体的业务场景下的知识哈。

作者回复: 多谢支持! 这门课会介绍到的推荐业务场景下的知识, 希望能帮助到你。



Geek_663208

2020-09-23

王老师的粉丝报道，常看您知乎，刷了两遍深度学习推荐系统，看到来极客开讲了果断订阅，看了下课程前三讲及目录，感觉肯定是一门干货满满的课，就是规划的三十三讲是不是有点太短了些哈，能否考虑多讲下应用深度学习模型过程中碰到的调优相关问题，想多了解一下实际场景中的坑及解决的相关细节，因为身边有千万级日活的组线上还是跑的LR+GBDT（虽尝试过很多深度学习模型但效果没提升）。

展开 ∨

作者回复: 因为考虑到不同基础的同学，所以有些点不会太深入进去。基本是以深度学习推荐系统主要模块的实现为线索开设的课程。

你说的实际场景下的坑和调优，确实是非常重要的经验，但考虑到各家的数据、环境、模型和数据的更新方式、业务场景可能都不一样，有时候一个细节就可能影响效果的成败，所以并不都可以提炼出来。

在课程中我也会穿插介绍一些实践经验，希望能让你大家收益。

