

Welcome

MSRA Researcher新手，现阶段从事多模态大模型预训练方向。对NLP感兴趣或对博客内容有任何疑问及意见建议的，欢迎评论或添加我微信。此外如果有需要内推的同学，也欢迎来骚扰我。联系方式详见concat页面。

前往GitHub (<https://github.com/Dodo>)

3月 20, 2019 (<http://www.pkudodo.com/2019/03/>)

学习规划 | 机器学习和NLP入门规划

作者 Dodo (<http://www.pkudodo.com/author/root/>) 在 (<http://www.pkudodo.com/2019/03/20/1-10/>)
NLP (<http://www.pkudodo.com/category/nlp/>),
机器学习 (<http://www.pkudodo.com/category/%e6%9c%ba%e5%99%a8%e5%ad%a6%e4%b9%a0/>) 标签
NLP (<http://www.pkudodo.com/tag/nlp/>),
机器学习 (<http://www.pkudodo.com/tag/%e6%9c%ba%e5%99%a8%e5%ad%a6%e4%b9%a0/>),
规划 (<http://www.pkudodo.com/tag/%e8%a7%84%e5%88%92/>)

阅读数: 13,760

每天来博客的人不太多，但稀稀拉拉也有一些固定的人群。自从上次发表了那篇面试的博客以后，陆陆续续就有一些同学来问我机器学习及NLP的学习路线，emm....因为有些问题有一些重复性，所以还是打算写一篇博客，把我的学习规划写出来。但是每个人都是独一无二的，我的计划可能对你有用，对他没有，也可能对别人有用，对你没用。需要自己去斟酌一下。

然后本文写的东西都是适合以后专门吃这碗饭的技术人员，不太合适其他方向需要使用机器学习技术的同学，比如说生物学的同学如果在数据分析上需要去学一些机器学习的方法来分析，就不太适合我这种学习方式了，没有必要去学那么深。

机器学习部分：

1.关于书籍

很多同学看的第一本这方向的书籍可能是西瓜书，包括我也是。但是这本书还是不建议大家看，它详细介绍了机器学习的各个方面，也就注定了每个方向都讲不深，有点泛泛而谈。还是强烈推荐李航的《统计学习方法》，书中有十来个算法，已经涵盖了机器学习的各方面了。

另外还有一本《机器学习实战》，封面印象里是一个农夫拿着锄头的，这本书只讲代码，理论讲得很浅。

一本理论，一本实践，两本书配合刚刚好。

2.关于视频

吴恩达的CS229比较强推，这个视频配合《统计学习方法》非常舒适。

其他视频像台湾有一个机器学习基石什么的课，评价也挺好，我舍友看的这个，他的意思是讲得比吴恩达好。但是如果让我推荐，我还是推荐大家看吴恩达的。

另外在B站上白板推公式系列 (https://space.bilibili.com/97068901?spm_id_from=333.788.b_765f7570696e666f.1), 讲得也很好, 我当时是挑着看的, 看完手上的东西还是理解不了, 可以看看这个视频。

3.关于学习方式

先看吴恩达的cs229, 在最初阶段方面入门, 同时一定要记笔记。期间《统计学习方法》等书可以不用看, 先把视频看完, 笔记记完。如果在视频中有看不懂的地方, 可以找一些博客看一看。看视频阶段最好在10到20天搞定, 最晚一个月也要看完了。

看完以后开始看《统计学习方法》, 一个一个模型看下去, 每次看之前可以先过一遍吴恩达视频记的笔记, 重新捡起来吴恩达的思路。这本书比较难啃, 但是一定啃下去。很多时候主要问题在不知道这个公式是怎么从上个公式推过来的, 多思考, 实在想不出来的, 网上找博客(比如说我的博客, 羞涩)。一定一定要跟着书把所有的公式推出来。

《统计学习方法》一个模型看完以后, 好好再回过头来看他给每个部分起的小标题, 再分析为什么书中这些小部分是顺序是这样的。你会慢慢理解推导一个算法的思路, 这个很重要。

至于一个模型怎么算初步学到位了, 我认为你学完以后, 可以假装边上有个啥也不懂的学弟学妹, 自己一点一点给他们讲, 给他们推导。如果整个过程都能讲清楚, 说明初步掌握了, 这时候已经打败了大部分人了。(最好是可以写博客)

在全书都看完以后, 给模型写算法, 可以看《机器学习实战》上的代码来知道机器学习算法应该是怎么写的, numpy是怎么用的, 但是一定不要照着书本写代码。对每个模型的公式充分理解, 在纸张写下来大概思路, 确定以后就开始写。(其实我觉得我的模型实现代码 (https://github.com/Dod-o/Statistical-Learning-Method_Code) 比这本书写得好一点....羞)

如果上面的部分都照着做完了, 我觉得你已经比大部分人优秀了。

之后在出去找工作前, 重新把《统计学习方法》过三遍, 第一遍照着书本推公式, 把思路捡回来。过几天第二遍, 抛开书本推公式, 如果有遗忘的地方看看书, 再好好想想你忘记的这一步在整个过程中起了什么作用, 基于什么的考虑, 自己怎么能保证下一次不会忘记, 再把书合上继续推。再过几天第三遍, 这时候已经可以不用看书, 说起哪个模型, 就能直接手推了, 如果还做不到, 就第四遍、第五遍.....

NLP部分:

NLP目前分为了统计学习方法和深度学习方法两大部分, 建议先从深度学习方法学起, 容易出成果, 内心成就感会高一点。

1.关于书籍:

统计学习方法: 宗成庆 《统计自然语言处理》

深度学习方法: 《TensorFlow实战google深度学习框架》

在NLP的深度学习方面其实我至今没有找到一本系统的书, 所以只推荐了一本google的深度学习。

《统计自然语言处理》其实我也不太喜欢, 但是他有配套视频(下面会讲), 所以顺带推荐一下。(不过我基本上没看过)

2.关于视频

深度学习方法：

吴恩达的深度学习视频，在网易云课堂上有中文版，强烈推荐。一共好像5个章节，其中一个章节是图像处理的，可以不用看。其他的我感觉十天左右也能看完了吧。这个视频看完以后对TensorFlow和NLP已经可以形成一个初步的理解了。它每个章都有作业，一定要做（作业在github上可以搜到）。

同样，一定要记得做笔记。

至于广泛流传的CS224N，我不太喜欢，但不排除你会喜欢。（其实是我不太看得懂，感觉还有一些先修课我可能没修）

统计学习方法：

宗成庆有个自然语言处理的视频，一共是64讲，这个视频一定要吴恩达的视频看完，做完几个练手项目，不知道该做什么了的时候才开始看。或者说你什么时候觉得自己需要补这方面知识了，再看就可以了。

3.关于学习方式：

首先看吴恩达的视频，很通俗易懂也很舒服，它里面分为每周的学习计划，其实每周的视频一天就能看完，再花半天一天时间做作业。全部看完以后，去做一些文摘生成、文本相似度比较、文本分类、机器翻译这些任务。但是不用找个什么练手项目100个，然后一个个做过去什么的。（挺想推荐我自己的github的，但是我已经很久不更新....）。每次订个目标，比如说这次我要把rnn实践一遍，就找相关博客，多看几篇，然后再去干活。主要是也没个什么系统的书，这方面学习主要靠博客。把rnn、lstm、seq2seq、cnn这些主流的东西都实践一遍，我感觉也差不多了。

这个时候如果你已经有了很感兴趣的方向，比如说你对对话系统很感兴趣，那可以直接开始往这个方向深入研究了。宗成庆的视频你什么时候觉得需要看了，再看就可以了。

我尽力去不误导大家，但是回头看，文章里还是写了很多绝对的话。学习的方法不是绝对的，我的这些方法，你觉得还行，那不妨可以试一试。要觉得差点意思，笑一笑过去了也就算了。

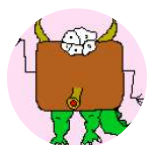
最后祝大家心想事成。



Dodo



💬 30条评论



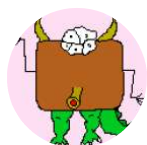
匿名

发布于10:03 上午 - 7月 2, 2023

博主还更新吗，统计学习方法后面还有挺多的

()

回复



匿名

发布于6:45 上午 - 1月 28, 2023

Hello pkudodo.com webmaster, You always provide helpful information.

()

回复



匿名

发布于4:54 下午 - 11月 11, 2022

大神！给你跪下了，把大哥你的博客基本看完了，学到了很多东西！

()

回复



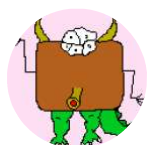
匿名

发布于11:12 下午 - 9月 9, 2022

bunadisisj nsjjsjsjsismizjzjjz zumzsert

()

回复



匿名

发布于3:33 下午 - 7月 27, 2022

特别感谢

()

回复



**匿名****发布于1:59 下午 - 6月 29, 2022**

非常感谢!

()

回复

**匿名****发布于10:12 上午 - 11月 23, 2021**

g7mlgzs

()

回复

**匿名****发布于10:04 上午 - 5月 6, 2021**

感谢大佬

()

回复

**匿名****发布于5:25 下午 - 10月 27, 2020**

多谢大神推荐!

()

回复

**匿名****发布于9:47 上午 - 8月 5, 2020**

谢谢推荐

()

回复

**匿名****发布于6:07 下午 - 6月 29, 2020**

请问吴恩达的CS229课程 我在youtube上找了个英文的, 发现看起来有些吃力 请问你那里有没有中文的CS229

()



回复



匿名

发布于5:18 下午 - 6月 16, 2020

太强了，参考参考

()

回复



匿名

发布于4:01 下午 - 6月 2, 2020

请问cs229是08版的嘛，和18版区别大不大

()

回复



匿名

发布于9:55 下午 - 5月 20, 2020

准研究生前来取经，感谢！

()

回复



匿名

发布于2:04 下午 - 3月 29, 2020

谢谢指引，赞一个

()

回复



匿名

发布于11:07 上午 - 3月 17, 2020

正在往这条路上走的小白，向大佬学习

()

回复



**匿名****发布于5:42 下午 - 1月 19, 2020**

mark,作为一个NLP入门新手来说这篇文章太好了。

()

[回复](#)**匿名****发布于10:01 下午 - 12月 20, 2019**

发现一个小宝藏，收藏了收藏了，c端想转ai~

()

[回复](#)**匿名****发布于10:00 下午 - 7月 9, 2019**

感谢

()

[回复](#)**匿名****发布于9:29 下午 - 5月 23, 2019**

谢谢您的分享，follow了您的github. 希望未来能够达到您的高度！

()

[回复](#)**匿名****发布于7:15 上午 - 5月 16, 2019**

老哥本科是Inkj? 真是牛皮

()

[回复](#)**匿名****发布于9:32 上午 - 5月 7, 2019**

关于NLP的不更新了吗

()



回复



Dodo

发布于11:42 上午 - 5月 9, 2019

github吗? 可能...最近不会往下写了 事情太多了 blog的话会开始写一些NLP的论文解读

()

回复



匿名

发布于5:11 下午 - 4月 8, 2019

这个博客用什么搭建的?

()

回复



Dodo

发布于12:01 下午 - 4月 10, 2019

wordpress

()

回复



匿名

发布于12:57 上午 - 5月 14, 2019

什么主题?

()

回复



Dodo

发布于2:37 下午 - 5月 15, 2019

chilly

()



回复



匿名

发布于6:41 下午 - 4月 2, 2019

写的真实在，点赞!!!

()

回复



匿名

发布于9:56 下午 - 3月 28, 2019

为北京吴彦祖打 call

()

回复



匿名

发布于2:59 下午 - 3月 22, 2019

感谢!

()

回复

发表评论

内容(链接请去除http协议头，否则会被误判为垃圾评论)

设置显示昵称(选填)

发送信息

近期文章

- ✓ 论文|记忆网络之Memory Networks (<http://www.pkudodo.com/2019/06/14/1-13/>)
- ✓ 梳理|对话系统中的DST (<http://www.pkudodo.com/2019/06/09/1-12/>)
- ✓ 论文阅读| How Does Batch Normalization Help Optimization (<http://www.pkudodo.com/2019/05/23/1-11/>)
- ✓ 学习规划| 机器学习和NLP入门规划 (<http://www.pkudodo.com/2019/03/20/1-10/>)
- ✓ 面试体会| 微软、头条、滴滴、爱奇艺NLP面试感想 (<http://www.pkudodo.com/2019/03/10/1-9/>)

文章归档

- ✓ 2019年6月 (<http://www.pkudodo.com/2019/06/>)
- ✓ 2019年5月 (<http://www.pkudodo.com/2019/05/>)
- ✓ 2019年3月 (<http://www.pkudodo.com/2019/03/>)
- ✓ 2018年12月 (<http://www.pkudodo.com/2018/12/>)
- ✓ 2018年11月 (<http://www.pkudodo.com/2018/11/>)
- ✓ 2018年10月 (<http://www.pkudodo.com/2018/10/>)
- ✓ 2016年9月 (<http://www.pkudodo.com/2016/09/>)



标签

DST (<http://www.pkudodo.com/tag/dst/>)

K近邻 (<http://www.pkudodo.com/tag/k%E8%BF%91%E9%82%BB/>)

LSI (<http://www.pkudodo.com/tag/lsi/>)

memory network (<http://www.pkudodo.com/tag/memory-network/>)

MQTT (<http://www.pkudodo.com/tag/mqtt/>)

NLP (<http://www.pkudodo.com/tag/nlp/>)

s3c2440 (<http://www.pkudodo.com/tag/s3c2440/>)

SLU (<http://www.pkudodo.com/tag/slu/>)

SVM (<http://www.pkudodo.com/tag/svm/>)

VSM (<http://www.pkudodo.com/tag/vsm/>)

体会 (<http://www.pkudodo.com/tag/%E4%BD%93%E4%BC%9A/>)

决策树 (<http://www.pkudodo.com/tag/%E5%86%B3%E7%AD%96%E6%A0%91/>)

分类器 (<http://www.pkudodo.com/tag/%E5%88%86%E7%B1%BB%E5%99%A8/>)

实现 (<http://www.pkudodo.com/tag/%E5%AE%9E%E7%8E%B0/>)

对话系统 (<http://www.pkudodo.com/tag/%E5%AF%B9%E8%AF%9D%E7%B3%BB%E7%BB%9F/>)

感想 (<http://www.pkudodo.com/tag/%E6%84%9F%E6%83%B3/>)

感知机 (<http://www.pkudodo.com/tag/%E6%84%9F%E7%9F%A5%E6%9C%BA/>)

支持向量机 (<http://www.pkudodo.com/tag/%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%9C%BA/>)

文章相似度 (<http://www.pkudodo.com/tag/%E6%96%87%E7%AB%A0%E7%9B%B8%E4%BC%BC%E5%BA%A6/>)

最大熵 (<http://www.pkudodo.com/tag/%E6%9C%80%E5%A4%A7%E7%86%B5/>)

朴素贝叶斯 (<http://www.pkudodo.com/tag/%E6%9C%B4%E7%B4%A0%E8%B4%9D%E5%8F%B6%E6%96%AF/>)

机器学习 (<http://www.pkudodo.com/tag/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/>)

爬虫 (<http://www.pkudodo.com/tag/%E7%88%AC%E8%99%AB/>)

统计学习方法 (<http://www.pkudodo.com/tag/%E7%BB%9F%E8%AE%A1%E5%AD%A6%E4%B9%A0%E6%96%B9%E>)



统计方法学习 (<http://www.pkudodo.com/tag/%e7%bb%9f%e8%ae%a1%e6%96%b9%e6%b3%95%e5%ad%a6%e>)

综述 (<http://www.pkudodo.com/tag/%e7%bb%bc%e8%bf%b0/>)

网易云歌单 (<http://www.pkudodo.com/tag/%e7%bd%91%e6%98%93%e4%ba%91%e6%ad%8c%e5%8d%95/>)

规划 (<http://www.pkudodo.com/tag/%e8%a7%84%e5%88%92/>)

详解 (<http://www.pkudodo.com/tag/%e8%af%a6%e8%a7%a3/>)

逻辑回归 (<http://www.pkudodo.com/tag/%e9%80%bb%e8%be%91%e5%9b%9e%e5%bd%92/>)

逻辑斯蒂回归 (<http://www.pkudodo.com/tag/%e9%80%bb%e8%be%91%e6%96%af%e8%92%82%e5%9b%9e%e>)

面试 (<http://www.pkudodo.com/tag/%e9%9d%a2%e8%af%95/>)

powered by wordpress (<https://cn.wordpress.org>) | copyright © 2018-2023 | 京ICP备18056879
(<http://www.beian.miit.gov.cn>)

