



erica119120 昨天 15:15 | 只看该作者 ▶

1楼 电梯直达

本楼: 100% (8) 0% (0)
全局: 100% (100) 0% (0)

我的人缘 0

3 6 440
主题 帖子 积分

发消



分享帖子到朋友圈

2019(10-12月) [MachineLearningEng](#) 博士 全职@[GoogleFacebook, Amazon](#) - 猎头 - 技术电面 [Onsite](#) 校园招聘会 [视频面试](#) | [Other](#) | 在职跳槽

我之前在另一个帖子里面分享了, ML design 面试的解题思路总结, 大家反应很有用。最近又在实践中总结出来一套ML design答题模板, 成功的过了几个大厂的ML design面试。因此分享给大家。

求加米, 给我米不会减少你自己的米, 举手之劳, 互惠互利。

面试就是要在有限的时间里, 尽量把自己懂的东西都让面试官看到。你懂那么多模型理论, 怎么才能让面试官也觉得你很懂呢? 要有策略, 有重点。什么样的策略好用呢?

抓住核心, 兼顾深度和广度。

ML design的核心, 万变不离其宗, 本质都是train一个model来实现某个task, 比如prediction/ranking/classification。有经验的人都知道, 实际上给定一个问题, 好用/常用的solution基本上只有很少的几种。所以想要显得你懂很多, 不仅要深度上要cover这几种solution, 更要从广度上显示你有end-to-end的经验。具体怎么实行呢?

Step 1: 理清核心问题。

- 不是每个面试官都能用一目了然的方式提问。有的面试官水平差, 自己也理清问题的逻辑。遇到问题很模糊的时候, 要尽快理清核心问题。抽象出来, 可用信息/输入有哪些, 要求的输出是什么样的, 这是一个classification的问题, 还是regression, 还是relevance/matching/ranking? 理清核心问题, 就能判断需要train哪种类型的model, 整个pipeline就很容易flow out了。

Step 2: 理清核心问题后, 请白板画图, 最好是一个diagram。有前后逻辑关系的work flow最能展示你思维的广度。

- 在理清核心问题后, 具体分析model之前, 先把solution的大体框架在白板上画出来。目的是让整个讲解过程逻辑清晰。按照逻辑的先后关系, typical的解答逻辑包括这几大块: training/testing data, input representation, model, output, evaluation, optimization(parameter estimation)。我一般从model开始画, 一个框框摆在中间, 这是核心。然后画上游, 下游。在这里, 只要把框架搭好, 告诉面试官, 我要讲这些内容, 面试官有个心理准备, 就可以开始听你讲课了。

Step 3: 讨论model。为什么我用“讨论”这个词? 因为能seriously被考到design的人, 都不是entry level。对于更senior的人来说, 面试的最好氛围不是你问我答, 而是我把我知道的都讲给你听, 你看看还有什么想听的。所以你讲的过程中要和面试官互动。要看ta的反应, 哪里皱眉了, 哪里表情不轻松了, 你就要停下来, 问他Is there anywhere that you want me to talk more? 这给面试官一个机会表达自己, 也帮助你更好的address面试官的考点。

- Model方面, 针对task的类型, propose哪些model可用, 把你能想到的都name出来。选择2-3个常用的, 比较优劣, 然后选择一个大家常用的。不同的model, 输入输出可能不一样。所以决定了model, 其他的component就很自然的浮现了。这一步, 要在你的model框框里, 把关键的component列出来, 说明它们之间的关系。分析各个model的优劣, 可能需要在旁边额外画出model的visualization, 比如说到dnn, 你就画几层multi perceptron layer, 再顺便提一下SGD和ADAM。说到用logistic regression 做classification, 你就顺手写一下log likelihood, 显得你optimization也很懂。说到regularization, 你就写写L1 norm和L2 norm。显示你的深度, 主要就靠这一步。有时候面试官会告诉你ta想用的model, 你就按照ta的来, 你也可以在讲解完几个model的优劣后, 根据经验自己决定一个model。

以下内容需要积分高于 20 您已经可以浏览

Step 4: 输入输出。

- 前面一步把model定下来。根据不同的model, 解释一下input 和output的format。比如dnn就有one-hot encoding, 这种最好用上embedding, 顺便讲一下有什么好处。比如需要自己设计feature的, 就重点讲一下有哪些常用的feature。

到此, 这轮面试的核心你都cover住了, 可以得到60分。Step 5和step 6 是能区别ML**和ML老手的部分。如果你做了, 答的一般可以再拿20分, 答的好可以再拿40分。有ML经验的人在这两个部分, 一定要把握。

Step5: data。不讲data的ML design是没有灵魂的。

-从2个方面identify data: training + label, testing + ground truth。1. propose可用的data来源+data format。2. how to preprocess data -> make training data, how to build/create label, etc. 完成建模。根据具体的问题, data的solution可以非常creative。甚至Training data和testing有时候不一致。比如language model方面的问题: decide一个twitter post是什么语言? training 可能就用wikipedia, testing则可以收集user data或者platform-specific的数据, 这时候也需要指明testing如何get ground truth(testing label)。

Step6: evaluation

- evaluation很容易讲，重点在metrics。主要有三个部分，一个是ROC/AUC curve。第二个是domain specific metrics，比如广告就有CTR。第三个是confusion matrix，重点是从它引申出来precision/recall/accuracy等等对你的solution重要的metrics。

最后再说几个加分项：

- 熟练的讲解参数估计，能显示solid的数学背景。讲估计参数可以用哪些optimization的方法(MSE, loglikelihood+GD, SGD-training data太大量, ADAM-sparse input)，比较优劣.
- 解答逻辑的每个部分，尤其是你熟悉的方面，要自己主动讲，因为每个部分都很重要。别问面试官想不想听你说，除非他明确制止你讲(如果面试官说不用讲，你就失去了一个展示自己的机会)。正确的做法是，lead conversation，一边白板画框图，一边告知面试官我要讲XXX这几个部分。整个design讲完了以后，再问面试官：Is there anywhere that you feel I missed?

求加米，等米看贴，给我米不会减少你自己的米，举手之劳，互惠互利!

The last but not the least: 每一步都尽快和面试官确认，move on，不耽误时间。想要在45分钟之内把面试官讲的心服口服，这么多个component，每个只有很少的时间。

○ 评分

参与人数	43	大米	+175	理由	收起
	李冬冬	+ 2	给你点个赞!		
	zm45919745	+ 2	给你点个赞!		
	赵晓得木	+ 3	给你点个赞!		
	韩乾Chad	+ 2	欢迎分享你知道的情况，会给更多积分奖励!		
	Florence090	+ 1	给你点个赞!		
	taotao11	+ 2	给你点个赞!		
	Huang0220	+ 2	很有用的信息!		
	Mint92	+ 2	很有用的信息!		
	fancyfish	+ 2	很有用的信息!		
	racoona	+ 2	很有用的信息!		