


ML&DEV[7] | 所谓算法工程师的工程能力是什么意思

 机智的叉烧

北京科技大学 统计学硕士

关注他

61 人赞同了该文章

导读：大家好，我是机智的叉烧，这是我ML&DEV系列下的第7篇文章，本系列下主要谈和机器学习算法模型和工程有关的内容。算法工程师的工程能力在很多地方都有谈到，但是到底这个工程能力是啥，这次我来给大家谈谈我的理解。

更多文章欢迎关注：

我的专栏：数学·数据·计算机

我的公众号：CS的陋室

我经历过得2019春秋招已经看到一些现在算法领域学习和工作需求之间的gap，近期随着秋招的进行，这个所谓的“工程能力”的问题又被拖上了风口浪尖，那么，到底什么是“工程能力”，他对求职者有什么要求，这次我带大家来探讨一下。由于本人只是新人，只有一年的实习经验，有些问题可能理解的不是很清楚，欢迎大家提出意见，一起探讨。

概述

我的理解，简单一句话说完，其实就和计算机领域里面的软件工程的很多说法非常类似，所谓的工程能力，对于一个算法/策略需求，你能从无到有地建立好模型/设计好策略，并将其投入实际应用，包括后续的监控迭代的能力。

那么从算法工程师角度，它包括那些内容呢？

- 数据的操纵能力。
- 模型的设计与开发能力。
- 项目的部署和监控能力。

当然的，还有一些更高端的或者说更加难说清楚的要求，代码的可迭代性、可读性、可迁移性等。

数据操纵能力

我说的是数据操纵，而非操作，是因为这里面有两个层面的使用。

一个是，懂得数据的操作，能够规范化，整理成模型需要的格式，甚至包括写SQL，确认数据的来源等等，这个对于大部分做过实验、有过实践操作的人来说其实不是一件难事。说白了，这个层面解决的其实是可行性问题。

第二个层面，能够更高效的完成上面的功能，你需要有一定的算法常识，找到操作的空间和时间复杂度尽可能低的方法，从而让程序更加高效的运行，那种把CPU、内存、GPU之类的打爆的程序肯定不可取，一定要有这个观念。

模型的设计和开发能力

首先是模型设计，说到设计，其实就是要求你根据实际情况来选择甚至是设计合适的模型算法来解决问题，学术界讲究高准确度之类的说法其实在现实应用并不可取，要考虑很多现实问题。首先，不是所有模型都适用于任何数据规模的情况，有的模型就是对小规模的数据有用，然后，对于不同的数据质量，不同模型的耐受能力是不同的，有的模型要求数据的正确性很高，但是实际上你遇到的数据会有很多问题，另外诸如样本不平衡等原因，很多模型其实并不适用，最后还有一点，就是实际的可用性的问题，举个例子，在现实场景，用户对一些等待的耐受性很高，可以等10秒，但是有些场景，其实用户等3秒就等不下去了，这个时候即使你的模型很强，但是时间太长，其实也不可取，这是个业务层面的问题了。

然后是开发能力，最近在知乎看到一个这样的问题，感觉略无语吧：

学习深度学习如何避免成为程序员？

478 赞同 · 23 评论 · 回答

目前的现状，我只能说不可能，你是需要开发能力的，哪怕你是做科研，你也得会写几行python代码吧。但我这里是想说，大家要是想找算法岗工作，那就需要一定的开发能力了。什么叫做开发能力，我的观点是能把你的模型发布出去让用户去用，这个理解起来很简单，你不能把模型的代码给一个用户去跑吧，那你要打包，要上线，可能你需要一些前端或者应用端的帮助，但是至少你这个模型能作为一个服务上线，能被人调用，这就OK了。

一般地，现在的主要开发代码就是c++和Java，现在的人力市场，多会一个优势大很多，如果会一些高级开发，那就更好了。

项目的部署和监控

模型做完了，要上线，这里的上线，危险程度就会提升，所以过程也就会根本更加谨慎。

- 对于模型，是否有不可容忍的bad case。
- 上线流程是都规范，如果是大项目，你的模块的更新是否会对有关模块产生影响。
- AB实验，实验的目标、内容以及检测的内容。
- 检测线上是否有问题，检测内容。
- 线上有问题后，是否有可马上回滚的方案。
- 埋点，一方面确认你的程序正常运行（监控），另一方面记录必要的数据，方便日后还可能要使用（迭代）。



「真诚赞赏，手留余香」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

算法 算法工程师 机器学习

▲ 赞同 61 ▼ ● 1 条评论 ↗ 分享 ♥ 喜欢 ★ 收藏 📄 申请转载 ...



写评论... | 你关注的 微调 等 4 个人关注了作者

1 条评论

默认 时间



寸进

说得好

2020-10-29

● 回复 ● 赞

文章被以下专栏收录



CS的陋室

微信公众号同名专栏！不定期更新技术文章！

推荐阅读

ML&DEV[17] | 算法工程师的技术深度

导读：算法工程师的发展路线会是什么样的，如何提升才是一个算法工程师合理的发展方向，做了一些思考和总结。 更多文章欢迎关注：
我的专栏：数学 数据 计算机 我的公众号：CS的陋室开发...

机智的叉烧

发表于CS的陋室

ML&DEV[5] | 系统理解特征工程

导读：大家好，我是机智的叉烧，这是我ML&DEV系列下的第5篇文章，本系列下主要谈和机器学习算法工程有关的内容。本期讨论玄学但似乎非常有用的特征工程，告诉你特征工程是什么，具体有...

机智的叉烧

发表于CS的陋室

ML&DEV[2] | 机器学习开发技能入门线路

导读：大家好，我是机智的叉烧，这是我ML&DEV系列下的第2篇文章，本系列下主要谈和机器学习算法工程有关的内容。上回讲到了有关数学知识方面的入门线路，给出了我的一些经验和学习建议...

机智的叉烧

发表于CS的陋室

ML&DEV[1] | 机器学习数学基础入门线路

导读：大家好，我是机智的叉烧。这是我ML&DEV系列下的第1篇文章，本系列下主要谈和机器学习算法工程有关的内容。数学是机器学习的重要基础，我作为一名数学出身的同学，做机器学习之后...

机智的叉烧

发表于CS的陋室