

加油站 | Practice makes perfect!

2019-06-07 蔡元楠 来自北京

《大规模数据处理实战》



你好，我是蔡元楠。

端午节了，先祝你节日快乐。不知道正在读文章的你这几天是不是有什么娱乐活动。我最近去参加了湾区一年一度的“三俗”活动——摘樱桃。而且，还看了一个有意思的电影《疾速备战》（John Wick: Chapter 3-Parabellum），推荐给你。

说归正题，在专栏的更新过程中，对于“留言的处理”这一点我还是要先道个歉，前段时间因为一些个人原因比较忙，所以对于问题的回复没有做好。我会努力把答疑这件事重新做起来。既然有问题了，就是要解决掉才行。

专栏进程也已经过半了，我好像感觉到同学们的阅读热情都在降低，排除“养肥了再读”的情况，可能是因为很多人进入了学习的“倦怠期”。我想了下，正好又到了假期，也许是应该专门写一篇文章，来跟你“聊聊天”。毕竟对于绝大部分人来说，“学习”这件事情真的是“逆人性”的，总需要一些建议和鼓励。

学习的目的就是解决问题

Practice makes perfect

- 学习的目的就是解决问题
- 工作需要你解决沟通问题
- 工作需要你解决技术问题

说实话，我们都不是天生以学习为乐趣的“学神”，所以现在我们的学习目的往往就是集中在“解决问题”上。而事实上，“解决问题”这件事其实并不是我们从工作的时候才开始面对的。

还记得我们在校园时期受的最多的训练吗？那就是“做题”。枯燥无味的“做题”过程就是我们接受的最初的对于“解决问题”的训练。在这个“做题”过程中学会的解决问题的方法一定是你至今都在受用的。

只不过是不同阶段里，你需要解决的问题的主题会变化而已。小时候应付数学题，生活中要应对和男 / 女朋友吵架，工作中要应对定义设计技术系统，等等。

当然，我们要面对的这些问题的规模也一直在变化。就像我们小时候做的题：根据 a 和 b ，求 c 是多少？已知条件都有，正确解法只有一个。而成人世界的问题：怎样改善北京的交通？已知条件你不是很了解，解法也不止一个，而且很难验证。

我十分同意 Amazon 创始人贝索斯的一句话，“与其看 20 年后什么会变，不如看什么不会变”。在我们的人生中，不会变的就是“人要解决问题”这件事。

工作需要你解决沟通问题

Practice makes perfect

- 学习的目的就是解决问题
- 工作需要你解决沟通问题
- 工作需要你解决技术问题

其实学习方法有很多，如果套用机器学习的术语，我们在学生时期常用的解决问题方法是 Reinforcement Learning 的 Imitation Learning，也就是模仿学习。孔子所说的举一反三也是在说这个道理，通过做一个题，学会做相似的其他题。

当然，不只是具体的技术问题，与他人的沟通交流中也可以用上这样简单的道理。

举个例子吧，之前有一段时间里，一直困扰我的问题是，我主持的会议议程总是被人打断。

Let' s take a step back, ...

Before we talk about this, let' s talk about...

.....

而解决这个问题的方法和学生时期一样，就是模仿。模仿那些把会议组织好的人，学习他们做了什么，看他们是如何沟通的，再套用到自己身上加以练习。这里我自己总结了很多细节，比如提前一天再把 agenda 发给参加会的人，会议开始前把 agenda 投影上去，被人打断了再强势打断回来，等等。

That' s an important topic, let' s have a separate meeting.

Let' s put it offline.

We only have 20min left, let' s...

.....

前面说过，学习的目的是解决问题。知道学习的方法只是第一步，更重要的是你要把它用到自己的工作、生活中。

工作需要你解决技术问题

Practice makes perfect

- 学习的目的就是解决问题
- 工作需要你解决沟通问题
- 工作需要你解决技术问题

说到这儿，我不知道你是不是了解我的工作，我给你简单介绍一下吧。其实这方面挺简单的，你随意 Google 一下，都能找到有关 Google Brain 的介绍。

Google Brain 的雏形起源于一项斯坦福大学与 Google 的联合研究项目。

2011 年，Google 资深专家杰夫·迪恩（Jeff Dean）、研究员格雷·科拉多（Greg Corrado）和吴恩达（Andrew Ng）教授是这个小团队的最初三名成员。团队有多个主要研究方向，包括机器学习算法和技术、医疗健康、机器学习支撑计算机系统、机器人、自然语言理解、音乐艺术创作和知觉仿真等。

我现在的工作方向就是医疗健康，主要负责的是癌症相关的 AI 应用。

比如，能否把所有癌症相关的数据更有效地整理起来，让医生和研究者更方便地查询。这些数据非常多，包括诊断过程中的 CT、活检切片、基因序列、治疗过程中患者对于不同疗法的反应，等等。

再举个细点的例子吧。有研究者发现，免疫疗法对于癌症患者的有效程度更多的是取决于癌变组织周围的免疫细胞环境。在这一前提下，我们要解决的问题就是：能否用 AI 来分析癌变组织的免疫细胞环境？

看到这，你大概能理解了，拥有**大规模数据的整理、搜集、存储、分析能力**，是我们完成工作内容的先决条件。而我们日常要锻炼的、加强的也就是这些解决技术问题的能力。

当然，可能大公司和小团队侧重的方面并不相同。如果你在一个小团队里，这个小团队里更看重的是你是否对这个事业有兴趣。但是，大公司主要看重的还是员工**解决沟通问题**和**解决技术问题**的能力。

这一点如果体现在面试流程的情境中，你可以这样理解：站在面试流程制定者的角度，他们设计一些面试流程的目的只是为了判定这个人进公司后是不是能成为一个好的同事，帮助公司成功。

所以，在面试的过程中，你可以先给自己一个人设，就是把自己当作是“面试官的同事”。把面试问题的内容定义成一个需要解决的问题，一个需要你们去共同解决的问题。比如，面试题条件没给全的时候，完全不需要惊慌，他很有可能是故意的，你需要做的是去和面试官沟通我们这个项目需求。

这与真实的工作中一样，拿到一个项目后，你要马上开始死命地工作吗？不是的。你应该是要多去问为什么。为什么做这个项目？团队想要达到什么目标？想要短期的方案还是长期的方案？这样的心态摆好之后，剩下的就是你的技术问题了。

从看到学，是一件事；从学到用，是一件事；从用到会，又是另外一件事了。而解决技术问题方法只有一个：Practice makes perfect!

今天说了这么多，其实都是我自己的一些经验。那么对于“如何解决问题”这一点，你可以分享一下你的经历吗？

© 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。

精选留言 (15)



—

2019-06-08

给老师正视答疑不足的心态并要认真解决的态度点赞！对了，我靠着老师的课程找到了心仪的大数据方向的实习工作，在这里也向老师表达衷心的感谢！

作者回复：恭喜你，加油。



👍 9



WL

2019-06-10

我感觉这个专栏可能比较适合有大数据开发基础的同学吧，像我这样没有大数据开发基础的开每篇文章都是走马观花，见见世面的感觉，感觉对当前的工作没啥实质性的帮助，如果能有专门的视频课程是入门spark, beam的，再回过头来看可能会感觉不一样。



👍 7



Geek_fredW

2019-06-07

棒！

知识和具体的技术是成长的必要条件。另一个必要条件就是这些“聊聊天”的经验、认知分享。

在我心里，解决问题最好的角度是：居高临下，通览全貌。当知道自己当前处在问题中的什么位置，往往至少会有一些进一步的思路。所以通常情况下我喜欢尽可能搞清楚问题的全貌。但实际上经常也会遇到没条件实现通览全貌的情况。这时另一个有效的角度，我自己的体会就是找到一个切入点，先试试。

前者是从上往下，从总到分的轨迹的话，后者就是反过来。

学习是一个艰苦的过程，现在也不是天天都有充足的时间学习。我就是这样的情况，默默学习。一些知识展开后工作量不少，再练习练习，其实花费的时间远远大于个把小时。所以我想所谓的学习热情减退，是一个现象，但它不是原因。相信有相当一部分人会坚持到底。

课程不错，我觉得我很幸运。棒！

作者回复：谢谢，一起加油。



👍 7



shawn

2019-06-10

学习新知识需要知道它解决了什么问题，是怎么解决的，如果没有它还有没有别的方法，如果还有别的方法那么可以对比一下之间相同和不同点

作者回复：✓



👍 4



kylin

2019-06-07

在一家外包公司做大数据开发，说白了就是用hive进行简单的数据处理，聚合。而且这些数据也不清楚有什么用，都是按照指示工作，很盲目，不知道眼前路应该怎么走下去



👍 3



hallo128

2019-06-08

对啊，每次您提出问题后，都不解答。与读者完全没有交流，我们也感觉您没有收到我们的反馈有所调整。也就很懈怠。希望您能多看看留言，进行解答，谢谢。



👍 2



刘万里

2019-06-07

老师请教个问题，最近在做技术选型，犹豫flink和beam两个，就请教下beam支持cep吗？

作者回复: 谢谢提问！Beam作为一个架构思想和Flink还是不一样的。如果你选择Beam SDK来编写你的数据处理逻辑，底层一样可以选择使用Flink来作为数据处理引擎。关于CEP的问题，这个看你的具体场景了，我认为Beam的编程模型里是涵盖了的。



👍 2



张凯江

2019-06-07

发现问题，分析问题，解决问题
说来容易，实施起来很难



👍 2



aof

2019-06-07

很多问题想着简单，真正去实施的时候会出现各种意想不到的幺蛾子，这也就是为什么很多程序员在接到任务的时候，说自己在多长时间之内就能做完，但结果是，最后因为实际中的各种问题而做不完，或者做完了，却检查出很多bug等待



👍 1



Hank_Yan

2020-07-16

我的解决问题方法论：明确问题，了解现状，积极调研，充分讨论，编写代码，code review，总结归纳。写代码只是整个工作中很少的一部分，前后的整个路子很重要。



stone

2019-07-05

方法论不错。

作者回复: 🍷



xj_zh

2019-06-19

熟能生巧

作者回复: 没错



happiness

2019-06-13

以解决问题为切入点来进行专门的学习和锻炼，这个思路非常赞，有的放矢，方得要领。还有我个人有幸听过关于google brog的一些设计思路和历史，请问老师有没有相关的资料可以推荐吗？



hua168

2019-06-12

大神，我看AI很多都是高级的数学公式，那不不要先把大学的高数、离散数学重学一遍？！



我都忘记玩了😭



商君

2019-06-11

从会到精，又是另外一回事了

