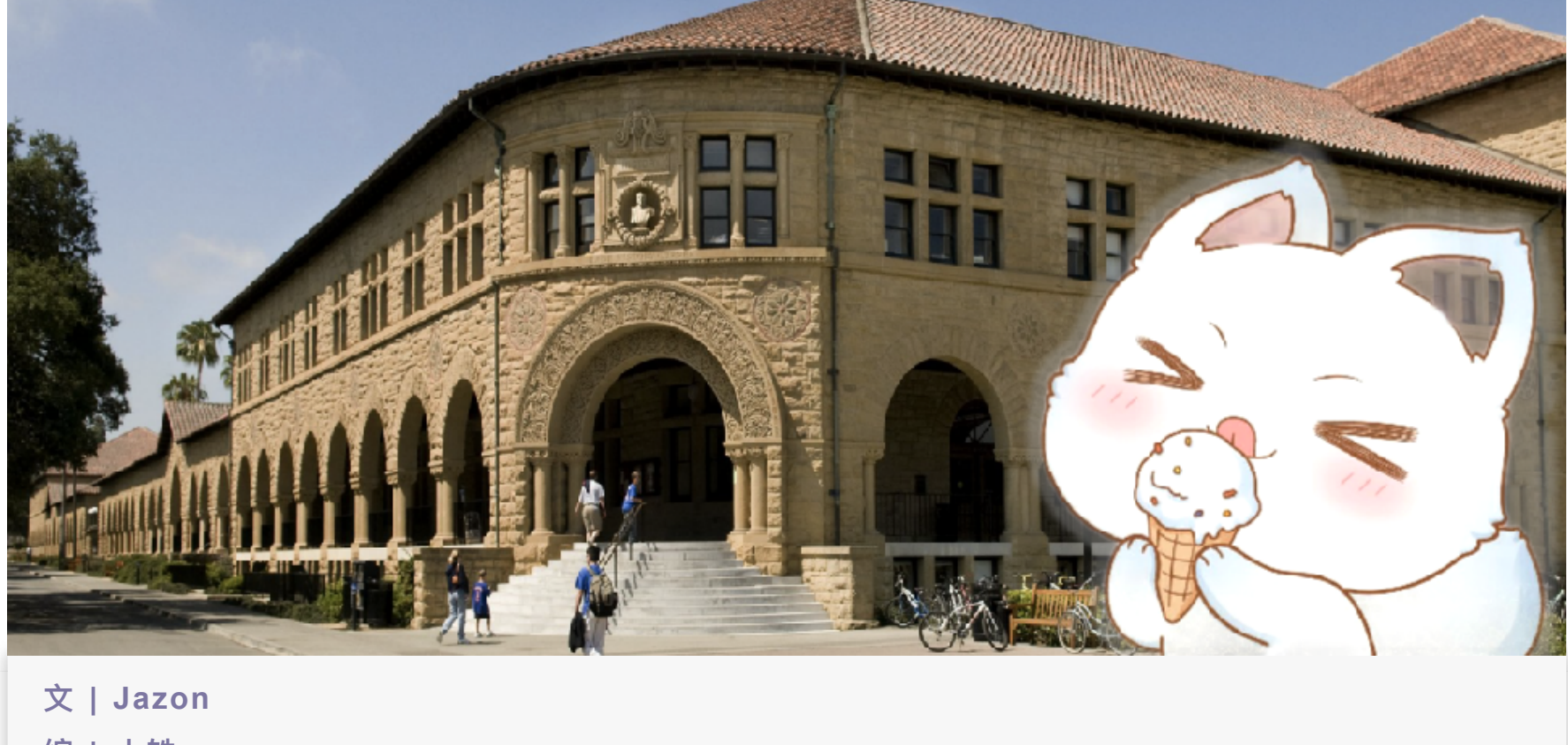


我在斯坦福做科研的碎碎念

原创 Jazon 夕小瑶的卖萌屋 2021-09-01 12:05



微信扫一扫
关注该公众号



文 | Jazon
编 | 小铁

大家好，我是 Jazon。作为 AI 科研工作者，我们的目光不该只聚焦在开发新技术、刷榜打 SoTA 上。学界如职场，还有很多技术之外的软实力需要培养。这篇文章里，我想以我个人的经历作为切入点，聊一聊与此有关的两个话题：

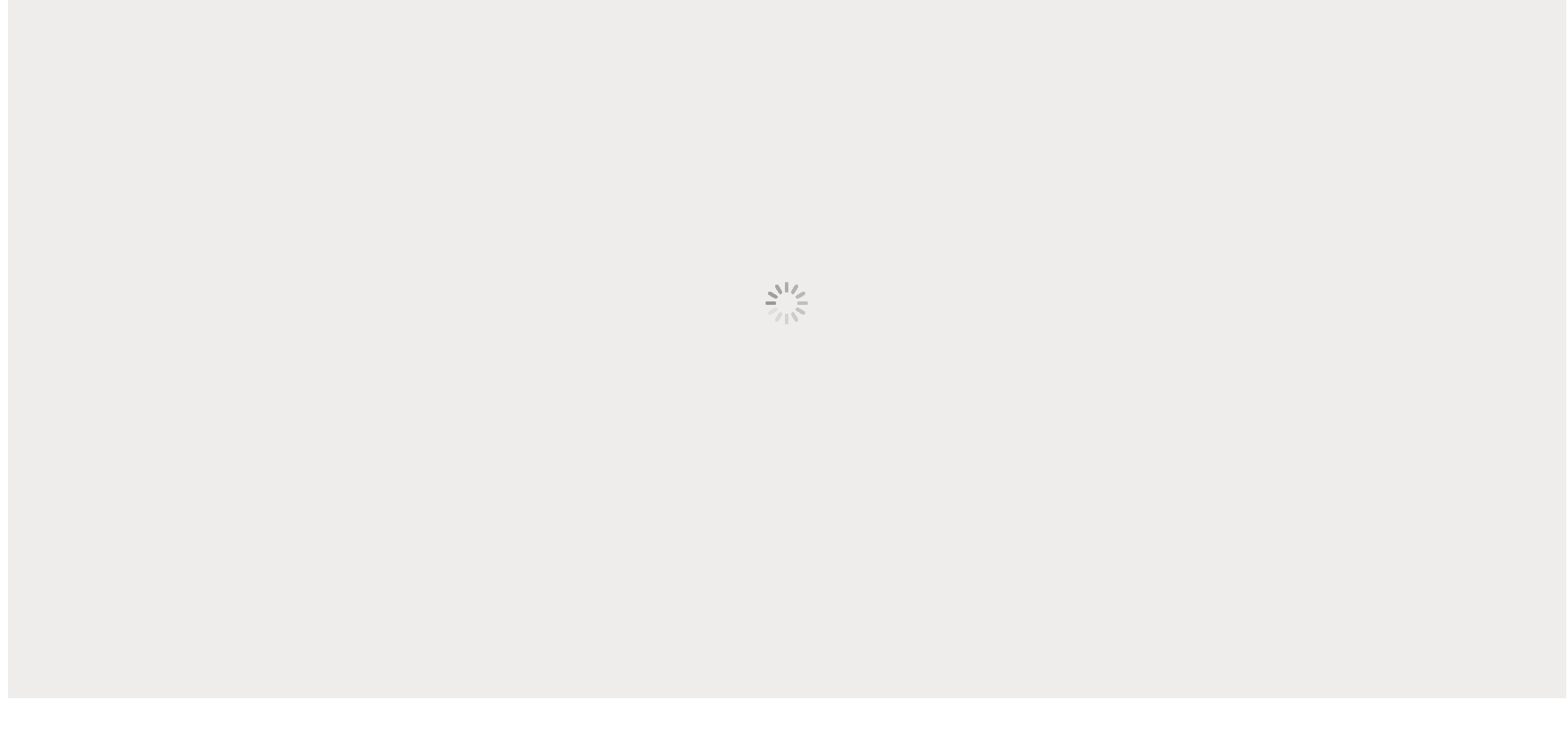
- 实验室文化；
- 团队合作下的代码质量。

背景介绍

在我之前的文章里提到，笔者现在在 Stanford 攻读计算机硕士项目；和中国的研究生不同，美国很多 CS 硕士项目（包括我们项目）中，做科研不是毕业的必须要求，也不需要要有固定的导师。一个学期里，可以在 lab 打工，也可以纯上课；如果在 lab 待得不喜欢，或者项目做完了，下学期可以选择换导师。

作为硕士生，寻找科研机会会有很多途径，比如直接联系老师或 PhD，或者有的 lab 有专门的申请表。另外，有些 project 的招人广告，会在系里以邮件的形式广播给所有同学。

2020年11月，我便通过这样一则招人广告，经过面试、测验，申进了 Stanford 商学院的 Social Impact Lab，之后直到研一结束（2021年6月），我在这里作为 RA（research assistant），做一个推荐系统方面的科研项目。Lab 的导师 Susan Athey 是经济学出身，美国科学院院士，24岁就博士毕业，维基百科上她的词条有13种语言，非常厉害。



关于我这次科研项目的技术细节，有兴趣的朋友可以访问我的个人网站查看（请参见文末“阅读原文”的链接）；本文主要是想吐槽一下技术以外的方面。

周报制度

刚入 lab，第一件让我头疼的事是lab的周报制度。在这里，我每周必须上报一个 **Progress Report**，里面主要分为几个模块：

1. 目标；
2. 职业发展；
3. 本周总结；
4. 下周计划。

其中，“目标”模块需要列出所有进行的 **project**，它们对应的 **deadline**，以及现在完成的进度。

“本周总结”要列出本周在每项工作上花了几个小时。

而“下周计划”要列出计划的任务，以及各自预计需要的小时数；每项 task 不能超过4小时，否则需要将其拆分成更小的 tasks。

刚入 lab 时，我对手头的任务还一头雾水，要我列出这么详细的计划，这让我很头疼。随着时间的流逝，我对项目渐渐熟悉，但写周报仍然是件很心累的事。

在我看来，**科研和其他工作不同，探索的方向有很大不确定性**，需要自由的思考空间来获取灵感。拿我的项目来说，我们有个大体的目标：提升某个平台上推荐系统的表现，细分下来有“用怎样的评分标准”、“冷启动怎么解决”等问题。但是更加细分的课题，不少都是在实验、阅读过程中自然出现的，无法提前预知。

另外，尝试量化每项任务的耗时，我觉得也不合理。写代码的工作本来就很难预测用时，更何况在科研的情境下，失败是家常便饭，经常会试了各种方法发现行不通后放弃，完成任务需要的时间就更不可控制。

每次的周报，我都得花大概1小时完成；如果没有按照规定的格式写，就会被 lab 的 HR 发邮件要求改正。半年下来我一共被四五次“建议修改”，都是些我觉得并没有必要的格式性问题，我也只是象征性地改一改应付一下，比如加一些实际上意义不大的“任务 deadline”。



▲ Stanford GSB

我们 lab 的领域之一是经济管理，理应深得领导力之道，但 lab 里这种“时间管理导向”的文化，让不喜欢刻意管理时间的我，很难适应。有一次我甚至吐槽，写这种周报是“血汗工厂”行为.....

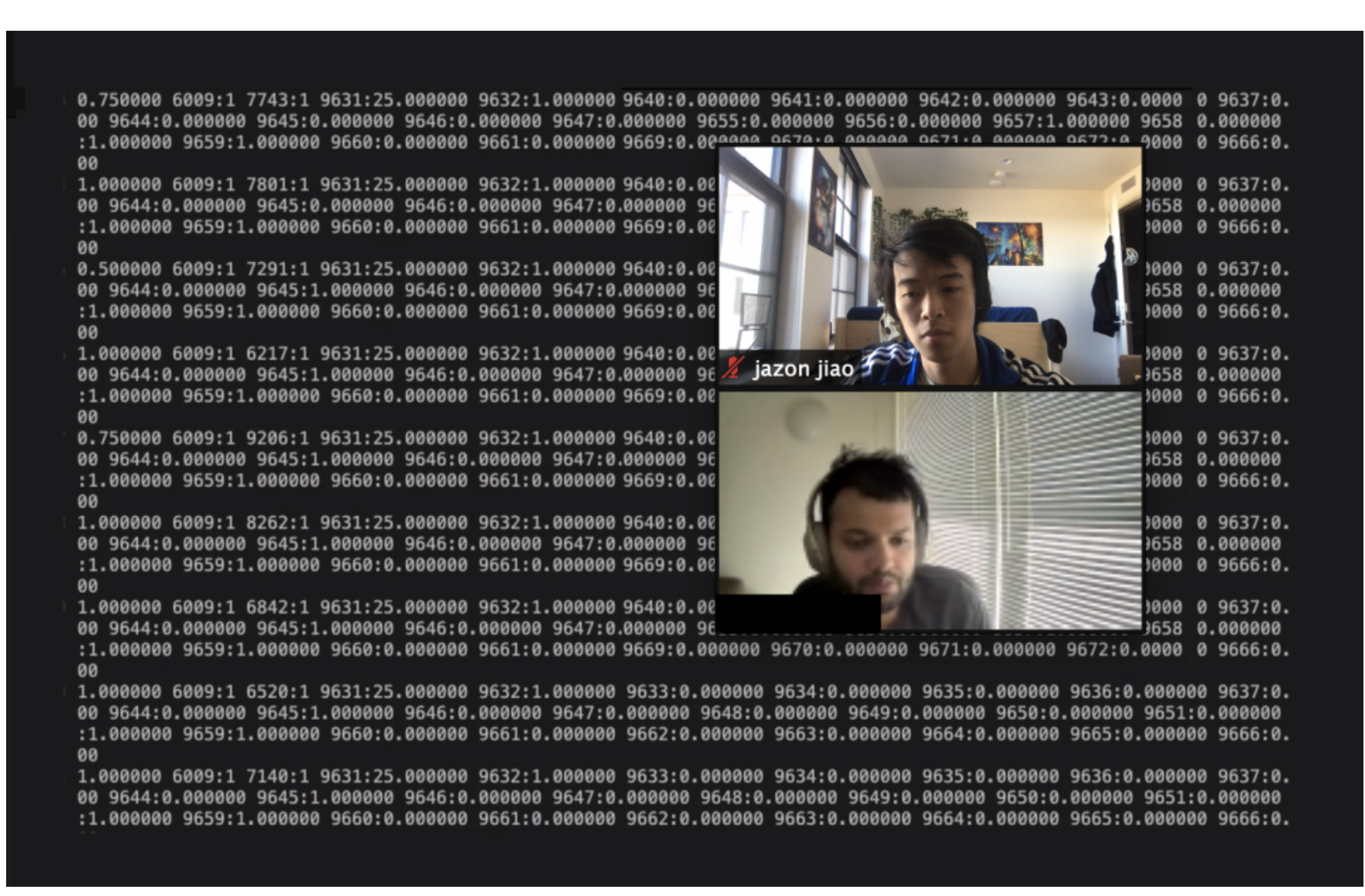
其实这个周报制度，是有不少好处的，有助于自己梳理工作进度、总结工作效率。另外周报里第2部分“职业发展”，要求列出近期的职业目标（计划去哪实习，或者何时开始求职），鼓励大家思考当下的工作对自己短期、长期职业规划是否有利，也很有意义。

只是我个人对周报的感受，更多还是负面的。这也反映了在某种程度上，我和 lab 的文化不太搭吧.....

代码质量

在不少互联网公司，都有对代码质量严格把关的机制。比如变量命名要遵循一定的标准，每个方法超过几十行要拆分；再比如有完善的 version control 系统，一段代码需要跑通测试、经过多人审核通过，才能部署使用。

但在很多实验室，大家写码普遍比较随意（如果你看过或用 AI 方面论文对应的 GitHub 代码库，可能对此会有同感）。如果是自己一时用，那没有问题；一旦是需要反复跑的，或者是需要别人合作的代码，随意写码的习惯就会带来问题。



4月的时候，我的 mentor 要求我在他写好的 pipeline 里加几个模块，跑一系列新实验。然而，在这个代码框架下，我连加一个 baseline 都非常困难。大致来说：

- 各个方法之间的关系错综复杂，如果出了问题很难 debug；
- 一个方法的输入，有时应是 pandas DataFrame，有时应是 PyTorch tensor，如果是 DataFrame，可能还需要含有特定名字的 column。这些都没有任何文档或者注释说明，加上由于是 python 写的，无法直接知道每个变量是什么样的数据结构；
- 一部分变量名写得不清楚，甚至可能有误导性，比如有个叫“input_file”的变量，实际代表的是 output 文件。

我和mentor反映了这套代码使用起来比较困难，他却好像不理解我的处境，觉得改这套代码是很简单的事。

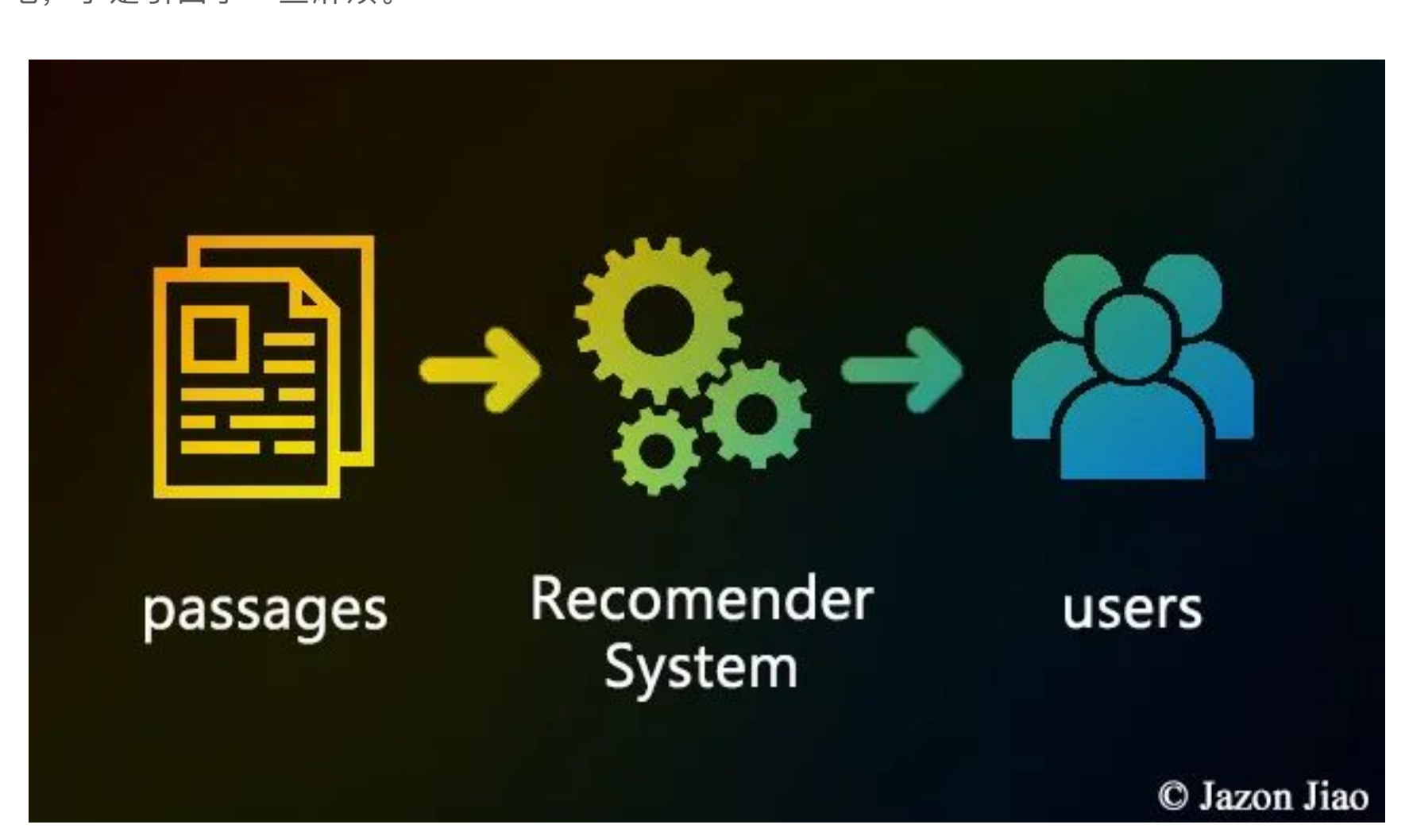
推荐系统小组里只有我们2个搞技术的，我自己说不动他，也找不到其他人可以帮忙；在产出实验数据的压力下，我只能违抗他的意愿，自己从头写了一套跑实验的 pipeline，从处理数据，到搭建模型，再到 train 和 test，最后可视化结果。

这样的好处是，我对代码的逻辑有完全的掌控，一来出了 bug 更容易 debug，二来如果要修改代码实现新功能，也能更快上手，灵活了很多。

当然，写自己的风险就是：我的代码相比原来别人写过的，可能有我注意不到的问题，或者代码的行为会有细微的、难以预测的差异，解决不好的话一样会卡进度。

果然，我的代码也出了问题。我犯了个低级错误，把 test set 当成 validation set 用；但这还是件小事，要命的是，我的 mentor 说，相比他花了很久搭好的 pipeline，我的跑实验代码“正确性尚未验证”，他就无法完全信服我跑出来的数据。这一点，后来直到我离职都没完全解决。

归根结底，软件架构这东西是需要一个牛人来搭建的，合作者去使用或加入模块应该是一件轻松的事情，不用管 pipeline 里其他的部分。但我在 lab 里的体验，就是处于与此相反的境地，于是引出了一些麻烦。



结语

这篇文章吐槽了很多，并不是说这次科研 project 的体验不好。实际上这是一次很充实的经历——探索未知课题的过程有很多乐趣，我的 mentor 和导师也很关心我，在技术上给了我很多指导，我非常感谢这次难得的锻炼机会。但我希望能总结一下感受不那么好的地方，希望自己，以及处在 AI 行业的读者你，可以从中学到一些道理~

萌屋作者：Jazon

来自南京，斯坦福MScS（计算机硕士）在读最年轻的中国人，预计2022年毕业。爱安静地探索宇宙的奥秘，也爱和朋友桌游、运动。梦想养猫，花花与三猫的视频平均每个看过20+遍。相信AI虽然有趣，但短期内在美国职场生存，还是要靠丰富的开发技能。

作品推荐

1. 人在斯坦福，刚上CS224n
2. 谢捷，人在斯坦福打SoTA



后台回复关键词【入群】

加入卖萌屋NLP/IR/Rec与求职讨论群

后台回复关键词【顶会】

获取ACL、CIKM等各大顶会论文集！

FOLLOW ME

STAR ME

阅读原文

喜欢此内容的人还喜欢

若被制裁，中国AI会雪崩吗？

夕小瑶的卖萌屋

