

《开源软件设计与开发》课程总结

姓名：郑宁远

学号：10165101164

1、开源和 Git 的理解

通过一个学期的学习，我认识到了开源社区存在的必要性。开源项目使得每一个用户都能参与到开发和维护中，这样才能建立健全的软件生态。许多大型的项目开发周期长，需要支持的功能多而杂，通过 git 这样的工具能有效地管理并让开发者熟悉项目的结构。新的参与者可以通过前人的某次 PR 来熟悉添加功能时所涉及的文件，这样就能逐步理解各模块的设计。同时，git 在本地开发的过程中也便于快速切换版本，只需付出少量学习成本就能节约大量额外劳动。

2、开源贡献

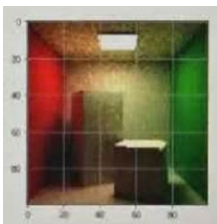
本学期我在亚马逊人工智能实验室参与了 Apache MXNet 的开发和维护，贡献了约 4000 余行代码，开了约 10 个 PR，发现兄弟项目 DGL 的一个安装 bug，提出了 issue 并得到了解决。这一经历让我认识到了大型软件开发中，团队内成员沟通的重要性。软件开发不仅仅是闷头写代码，很多时候充分的交流可以更快的发现和解决问题，节省大量精力。我的贡献在以下链接中可以查看：

<https://github.com/apache/incubator-mxnet/pulls?utf8=%E2%9C%93&q=author%3Avexilligera+>

我所在的小组负责开发 DeepNumpy，这是一个兼容 Numpy 支持自动求导的 MXNet 模块。我主要负责维护和开发其中的 linalg 线性代数库。

在工作中，我将 MXNet 中原有的一些算子（如 inv, cholesky 等）加入到了 DeepNumpy 接口中。同时，我修复了一个从 8 月份开始存在的 Windows 平台上内存访问越界的 bug，从而使框架支持了行列式的正向和反向运算。另外我实现了各类范数计算的算子，并且添加了大量的测试代码。我的代码使得其他人能够完成一些更高层的机器学习算法演示。

基于我们开发的框架，我自己也做了一些使用 demo。通过把路径追踪的并行运算转换成矩阵操作，利用 DeepNumpy 实现了一个可微的路径追踪渲染器。结果如下图所示。



然而，在工作中，我也发现了该项目大量的设计缺陷。在性能测试时，我们发现 DeepNumpy 普遍慢于普通 Numpy 五到十倍。这主要来源于函数调用时各类切换产生的额外开销。因此，我们社区接下来的一大任务是消除这些开销以提升性能，使其运行速度更接近原生 Numpy。另外，目前算子覆盖面虽然已较广，但有的操作写起来仍然不甚方便，一项长期目标是兼容更多的算子。

虽然作为开发者，但我认为 MXNet 的社区薄弱，用户少，开发者极稀缺，虽说有几家

大公司背书，但给我的感觉是钱砸了太少，只是用来撑门面的，已经远远落后于 Tensorflow 和 PyTorch，毕竟像各种框架这样的行业基本都是赢者通吃的，MXNet 马上就要没什么实际价值了，居然还搞 DeepNumpy 这种有的没的功能。建议有空不如把 API 统一一下，把大大小小的 bug 多修复一些。如今 MXNet 除大力支持 TVM 以外不具备明显优势，官方文档不完备。如果 MXNet 还想有活路，建议亚马逊多投入开发，否则靠几个实习生覆盖算子，人员流动大，核心代码维护者都跑了，连自己人都说难用迟早要完。以下为真实用户观点，我感到无奈与赞同：



徐文强

有阅历自成逼格。

1 人赞同了该回答

9102年都快完了，不支持boolean index，没有任何简单的方法workaround。简直就是ridiculous，不是被逼的绝对不会用它，建议mxnet要发展走上层路线，只要把老板们搞定了，下面的人哪怕恶心了也只能用。

----- update一下 -----

简单来说，就是文档太烂。想用的人都看不懂。

编辑于 2019-11-11

▲ 赞同 1



● 添加评论

➦ 分享

★ 收藏

♥ 喜欢

...

3、感想和反馈

通过将近一学期的工作和学习，我了解了开源软件的历史和开发流程，参与了开源项目的开发和维护。在这过程中，我掌握了 C++ 和 Python 的使用，了解了代码规范，熟悉了 Anaconda 环境的管理和安装，学会了使用 SSH 连接服务器，学会了用 VSCode 进行远程开发。更重要的是，我初步学会了 git 和 vim 的使用，这给我的生活带来了极大的便利。我今后打开 vim 再也不用重启电脑了。由于工作需要，我还学会了 mac 和 linux 系统的使用，并且能够在 mac 键盘和标准键盘之间熟练切换。同样由于工作需要，我学会了 PyTorch 和 Tensorflow 的使用，并且学会了在多机多卡上进行分布式训练，了解了 C++ CUDA 编程。同时我掌握了各种机器学习和深度学习算法，训练了各种 CV 和 NLP 模型，并且能够在此基础上加以改进试图提升 SOTA 了。这些经验给我今后的学习和工作带来了很大的帮助，为我深入学习打下了基础。

对于这门课程，我希望今后能邀请更多大型开源项目（如 Spark、Kubernetes、Tensorflow 等）的核心贡献者来传达他们的设计理念和未来路线图，给同学们牵线搭桥提供相应的机会来参与到这些项目的开发中。同时，希望能够和其他高校或企业实验室（如 BAIR、MPI、Google Brain、FAIR 等）合作，给同学们提供更多科研的机会，接触到更多的领域。

4、参考文献

无