

开源软件期末报告

王夕 10165102211

背景

我在本学期参加了开源软件 **Apache MXNet** 的开发，我总共提交了20余个 Pull Request,具体可以在 [github](#)，所有contribution代码行数超过5000行。我的工作内容主要集中在 `Deep Numpy`，一个以 **MXNet** 为引擎的提供类似 **Numpy API** 的深度学习框架，在这个框架中，我主要负责 `numpy.random` 模块的设计，实现和维护。除此之外，我也积极参与code review和issue处理，并与社区成员就许多设计问题展开了深入，友好的交流，详细内容可以在 [issue section](#) 查看，接下来，我会挑选几个PR进行介绍：

工作介绍

Rejection Sampling on GPU

在这个PR中，我实现了伽玛分布的采样，但这个采样有个实现难度，Gamma Distribution但采样需要用拒绝采样，这是一种需要使用到无限循环的采样方式，这就导致了在GPU上的潜在性能问题。为了解决这个问题，我首先在开了一个RFC issue，与社区成员展开了讨论：[15928](#)，在与擅长GPU编程的开发者(ptrendx)和random module的其他开发者(yzhiliu)进行讨论之后并进行POC后，我说服了社区成员我的思路，并将其实现。

Differentiable Normal Sampling

`Stochastic Gradient Estimation` 是一个历史非常悠久的问题，目前一种非常流行的方法是使用 `Pathwise Gradient`，我觉得这会是一个非常有用的特性，于是我首先在issue section 提出了RFC，并对我实现的 `np.random.normal` 进行改进，加入 `backward support`，在实现prototype之后，我提交了PR。然而，PR并没有立即被merge，一位社区成员提出：我们应该先有个scope，作为回应，我提出这个特性可以作为**MXNet 2.0**一个模块的基础架构，并成功说服了他，得到了approve。

Gluon.distribution

我在阅读学习社区置顶的 **MXNet 2.0** roadmap时，对其中的一个模块，`Gluon.distribution` 产生了浓厚的兴趣。在进行过相关背景调查，文献阅读后，我主动联系了负责人，并提交了一份设计文档，负责人在阅读完我的设计文档后，觉得可行，并与我在每周展开meeting，我在meeting中会向他汇报进度，并与他交流在工程中遇到的难点及新的想法，这个project将在2020年第一季度发布。

感想

经历了这么个月在开源社区的磨练，我的个人能力收获了巨大的提升：

在社区成员的监督下，我改掉了以前使用拼音给变量名命名的坏习惯，CI中的sanity check督促我写程序必须要有缩进。同时，我还掌握了如何使用礼貌的英语与社区成员沟通，已经能看懂如 `LGTM`，`np` 等缩写和一些专业术语，如 *infrastructure*(基础架构)。更重要的是，我还学会了 `git` 和 `google` 的使用，这两样工具对成为开源社区贡献者是必不可少的，在参与**MXNet** 开发之前，我从来没有接触过版本管理系统，我都是依靠建立多个副本并修改文件名来实现多版本控制的，我在遇到问题时，也往往求助于**CSDN**和百度知道，而不是更加强大的谷歌，为此，我还专门充值了288元购买了一年的VPN(虚拟私人网络)。除此之外，由于 `MXNet` 是一款支持 `Linux` 的深度学习框架，而我的电脑是 **Windows 10** 的，所以我特地购买了《鸟哥的Linux私房菜》来熟悉**ubuntu**这一款开源操作系统的使用，我学习了各种命令行命令和工具，包括 `make,cmake,ninja,cpplint,ls,vim`，这些复杂的命令行操作起初让我一个**CodeBlocks**用户感觉到非常的不适应，但后来我成功克服了困难，是开源社区的力量鼓动我去踏出自己的舒适圈，学习新的知识，认识新的人