

# ThunderGBM: 快成一道闪电的梯度提升决策树

想在 GPU 上使用使用闪电般快速的提升方法？了解这个库就好了。在很多任务上，它都比 LightGBM 和 XGBoost 快。

尽管近年来神经网络复兴并大为流行，但提升算法在训练样本量有限、所需训练时间较短、缺乏调参知识等场景依然有其不可或缺的优势。目前代表性的提升方法有 CatBoost、Light GBM 和 XGBoost 等，本文介绍一项新的开源工作，它构建了另一种基于 GPU 的极速梯度提升决策树和随机森林算法。

项目地址：<https://github.com/Xtra-Computing/thundergbm>

那么我们为什么加速 GBDT 和随机森林？2017 年，一份 Kaggle 调查显示，数据挖掘和机器学习从业者中分别有 50%、46% 和 24% 的人使用决策树、随机森林和 GBM。GBDT 和随机森林经常被用来创建当前最佳的数据科学解决方案，这就要求我们能使用 GPU 在大型数据集上完成高效训练。



两棵决策树的集成方法，选自 [XGBoost](#) 文档。

尽管 XGBoost 等库已经支持 GPU 了，但毕竟不是一开始就为 GPU 而设计的，因此在优化和加速上会有一些瑕疵。而 ThunderGBM 旨在帮助用户轻松高效地应用 GBDT 和随机森林来解决问题，它可以利用 GPU 完成高效训练。



对于 GPU 而言，ThunderGBM 在很多任务上都比其它几个库要快。

ThunderGBM 的主要特征如下：

- 通常是其它库的 10 倍。
- 支持 Python (scikit-learn) 接口。
- 支持操作系统 Linux。
- 支持分类、回归和排序。



ThunderGBM 预测和训练的整体流程。

ThunderGBM 主要作者包括新加坡国立大学的 Zeyi Wen 和 Qinbin Li、华南理工大学的 Jiashuai Shi 等，指导教师为 NUS 的 Bingsheng He。

## 入门指南

ThunderGBM 要求开发环境满足 cmake 2.8 或更高版本；对于 Linux，使用 gcc 4.8 或更高版本；C++boost；CUDA 8 或更高版本。

下载：

```
git clone https://github.com/zeyiwen/thundergbm.git
cd thundergbm
# under the directory of thundergbm
git submodule init cub && git submodule update
```

在 Linux 上构建：

```
#under the directory of thundergbm
mkdir build && cd build && cmake .. && make -j
```

快速测试：

```
./bin/thundergbm-train ../dataset/machine.conf
./bin/thundergbm-predict ../dataset/machine.conf
```

成功运行后, 你将看到  $RMSE = 0.489562$ 。

### 相关研究

如果读者对实现的技术及模型细节感兴趣, 可以查阅原论文:

论文地址: <https://www.comp.nus.edu.sg/~wenzy/papers/thundergbm.pdf>

其它相关文献:

- 论文: Efficient Gradient Boosted Decision Tree Training on GPUs
- 作者: Zeyi Wen, Bingsheng He, Kotagiri Ramamohanarao, Shengliang Lu, and Jiashuai Shi
- 地址: <https://www.comp.nus.edu.sg/~hebs/pub/IPDPS18-GPUGBDT.pdf>

工程 决策树 梯度提升 GPU ThunderGBM

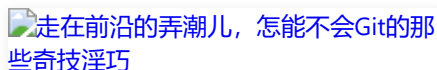
♡ 4 1 分享

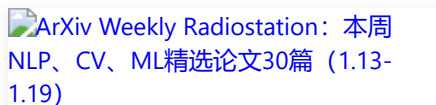
### 相关数据

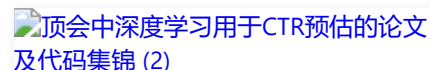
- [提升算法](#) • 技术 ∨
- [梯度提升](#) • 技术 ∨
- [神经网络](#) • 技术 ∨
- [随机森林](#) • 技术 ∨
- [数据挖掘](#) • 技术 ∨

展开全部数据 ∨

### 推荐文章

 [走在前沿的弄潮儿, 怎能不会Git的那些奇技淫巧](#)

 [ArXiv Weekly Radiostation: 本周 NLP、CV、ML精选论文30篇 \(1.13-1.19\)](#)

 [顶会中深度学习用于CTR预估的论文及代码集锦 \(2\)](#)

走在前沿的弄潮儿，怎能不会 Git的那些奇技淫巧

 机器之心


❤️ 3

ArXiv Weekly Radiostation: 本周NLP、CV、ML精选论文30篇...

Radiostation

❤️

顶会中深度学习用于CTR预估的论文及代码集锦 (2)

 机器学习blog

❤️ 1

**登录** 后评论



文祥

💬 ❤️

什么样的GPU支持呢?



全球人工智能信息服务

友情链接: [Synced Global](#) [机器之心 Medium](#) [博客](#)

[PaperWeekly](#) [动脉网](#) [艾耕科技](#)

[关于我们](#) [服务条款](#)



©2021 机器之心 (北京) 科技有限公司 京ICP备14017335号-2