第五次作业

**一、组内project的相关内容：**

**题目：**基于CNN的手写数字图像识别的并行设计与评测

**题目来源：**服务于课题组后期研究优化，选取信息主要由各结构和功能简单的神经元分散保存和处理、具有天然分布式存储和并行处理特性的卷积神经网络算法进行并行化评测，暂定数据集为经典的手写识别体MINIST数据集。

**最终目标：**首先在天河平台上进行多进程的并行测试；后期尝试引入GPU加速计算过程；关于并行策略，目前拟采用训练样本的数据并行。有余力的情况下，可以考虑关于特征图、神经元或是权值的并行性是否可行。最终计算各种情况下的加速比、并行效率、可扩展性评估，进而进行对比分析。

**实验条件及环境：**天河超算平台；GPU

**实施计划：**

|  |  |
| --- | --- |
| 时间 | 拟实现的任务 |
| 14周 | 复现串行CNN识别手写体的代码，并将其移植到天河平台运行试验测试。 |
| 15周 | 主要对算法进行数据并行改进。 |
| 16周 | 对不同数据规模训练样本在单节点、多节点的测试、记录并分析。 |
| 17周 | 引入GPU加速计算过程，完成代码的修改。 |
| 18周 | 同样对不同数据规模的训练样本进行测试并记录。 |
| 19周 | 对使用GPU前后的两种实验进行比对分析。 |

**第一周目标：**

复现串行CNN识别手写体的代码，将其移植到天河平台运行试验测试并记录。

**二、本组project收到的问题：**

实验强度太大，非本课程重点。

**改进：**对课程项目要求的重点理解有失偏颇。直接从问题的系统解决思路出发，没有考虑到实现难度，考虑到CUDA底层编程的复杂性，我们采用tensorflow实现。

**三、对其他感兴趣小组的问题及记录：**

**SpMV并行程序的可扩展性测试与分析：**采用稀疏矩阵和矩阵向量乘的实现机制上有什么区别？以及是否对符合特定特征的稀疏矩阵存储为某种特定的文件格式会有更好的性能？

**非嵌入式内存利用率、性能与高效能比**：功耗、（能量有效性）五秒记录一次每个节点部件实际功耗，根据功耗分解技术得到每个节点内部各个部件的实际功耗。功耗建模（数据采集拟合）怎么得到实测值进行评估，开发板接入功率计得到实际值。数据规模要达到多少（误差才能控制在百分之二）（实验中采用RAPL作为实测值精度是允许的）。

**共享内存/分布式内存结构下访存带宽测试与分析**：天河cluster，跨节点不能多线程，不共享内存。固定进程规模，不同的进程分布策略对性能的影响。什么是访存受限？memory latency会延长。什么样的输入集会造成访存受限？硬件决定的是memory bandwidth；特定程序执行的指标是memory traffic；每次访存的访问延迟是固定的，比如200cycle。

**典型科学计算程序的并行性能和功耗评估的评测与分析**：研究计算节点的能量有效性。总功耗不增加的情况下最大程度提升性能。计算密集型、访存密集型。功耗不变，增加节点数。

**并行应用通信行为特征评测与分析**：（tensorflow并行化后的通信开销）不同框架不同种类的开销均进行分析。