**基于Torch平台的神经网络压缩**

**研究与应用**

**项目计划变更说明书**

小组成员：

陈伟民

付强

曹进

李恬霖

# 项目计划变更情况

|  |  |  |  |
| --- | --- | --- | --- |
| 项目内容 | 初始计划 | 改进计划 | 变更原因 |
| 工作内容 | 1. 了解Torch架构和工作流程； 2. 分析Torch需求； 3. 压缩神经网络的编码； 4. 测试； 5. 其他软件管理工具学习 | 无 | - |
| 参加人员 | 陈伟民  付强  曹进  李恬霖 | 无 | - |
| 项目期限 | 本学期第二周至第十七周 | 无 | - |
| 项目分工 | 陈伟民：制作PPT，答辩；任务安排；制定交付文档的大纲；主要负责编码工作和测试工作  付强：答辩；管理GitHub；资料收集；主要负责非功能需求的编写和测试用例设计  曹进：答辩；计划制定与记录；主要负责制作各种用例图  李恬霖：记录与总结评审意见；环境搭建；资料收集；主要负责功能需求的编写和测试用例设计 | 陈伟民和李恬霖主要负责算法的研究和测试文档的撰写；  付强和曹进主要负责需求文档的撰写，并协同管理GitHub；  每名成员均承担一部分文档的撰写和编码的工作； | 充分发挥人员作用；协调各成员时间；分散每周任务压力； |
| 组织形式 | 1. 每周一次会议，讨论当前工作进展以及下一周的工作内容，讨论各自遇到的问题； 2. 建微信群，保持实时的交流； 3. 资源的共享通过GitHub和微信群文件进行实现； | 无 | - |
| 交付产品 | 1. 项目计划书 2. 项目进度表 3. 项目工作日志 4. 软件需求规格说明书 5. 需求检查单 6. 测试需求规格说明书 7. 软件测试分析报告 8. 软件测试评审报告 9. 项目工作量统计报告 10. 项目方法总结报告 11. 项目个人总结 | 1. 软件项目计划书 2. 需求规格说明书 3. 需求评审检查单 4. 实现方案改进说明 5. 测试需求规格说明书 6. 软件测试分析报告 7. 测试评审检查单 8. 软件测试评审报告 9. 项目统计分析报告 10. 配置管理总结报告 11. 计划变更分析报告 | 根据各实验要求调整 |
| 软件功能 | 网络剪枝 | 删除 | 未经严密设计，项目时间有限，完成算法编码难度较大 |
| 权重共享 | 删除 |
| 霍夫曼编码 | 删除 |
| - | 增加：权重二值化（随机投影矩阵与训练投影矩阵） | 算法相对简单，编码量符合实际时间允许情况 |
| - | 增加：不同求导方案（去sgn方法与Htanh方法） | 二值化后激活函数不连续，导致求导过程需要进行额外处理 |

# 项目执行情况

|  |  |  |
| --- | --- | --- |
| 项目内容 | | 执行情况 |
| 工作内容 | 了解Torch架构和工作流程 | 按计划执行；已完成 |
| 分析Torch需求 | 按计划执行；已完成 |
| 压缩神经网络的编码 | 按计划执行；已完成 |
| 测试 | 按计划执行；已完成 |
| 其他软件管理工具学习 | 按计划执行；已完成 |
| 参加人员 | | 按计划执行 |
| 项目期限 | | 按计划执行 |
| 项目分工 | | 按计划执行 |
| 组织形式 | | 按计划执行 |
| 交付产品 | | 按计划执行 |
| 软件功能 | 权重二值化 | 按计划执行；已实现 |
| 不同求导方案 | 按计划执行；已实现 |

# 计划变更类型

分工的变化一是为了权衡组内成员的特长，例如曹进有做用例图的经验；二是为了调配各成员的时间，减轻工作压力，因此曹进协助付强管理GitHub；三是为了让成员们都能够大体了解整个项目的内容和进度，提高编码能力，所以在文档撰写和扩展功能实现上让成员都有所参与。

项目扩展最初的计划是对Torch内部的层实现进行改进，改进的方向包括网络剪枝（移除不重要的连接）、权值共享、霍夫曼编码（利用有效权重的有偏分布）等。主要的工作重点在于算法的研究和对所有层的修改。

变更后，我们的计划改为对Torch的卷积层和全连接层（线性层）的计算进行修改。我们设计了权重二值化的算法，在原有实现的基础上，增加了投影矩阵，对二值化后不同的求导方案做了设计，并设计了基于受训练的投影矩阵的算法。项目变更后，我们确定了两个层作为修改的目标，工作重点在于编码，即对Torch中的源代码进行阅读并对其进行修改，不需要大量的时间进行算法的研究。

扩展项目类型由原来的算法研究，变更为现在的编码，并且降低了工作量，在后期的编码过程中，我们认为工作量是合适的。

# 变更原因

* 更好发挥组员特长，平衡工作压力，提升组员参与感。
* 对Torch实现的原理和细节把握不够强。
* 有限的时间内，很难完成大量算法的编码和测试。