计算机与通信工程学院硕士学位研究生中期考核表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学生姓名 | 李冠辰 | 学号 | M202120812 | 导师姓名 | 何杰 |
| 论文题目 | 基于结构重参数化的任务自适应模型自增长式训练研究 | | | | |
| 论文进展情况：  任务要求：  从任务自适应的角度出发，研究神经网络模型的自增长式训练。对于一个数据集（任务），以一个空的、几乎不含参数的模型作为初始模型，对模型进行训练；在训练的过程中，对模型的体量进行增长，包括宽度（低宽度层替换为高宽度层并进行参数填充，采用结构重参数化技术实现来进一步提升模型效率）和深度（插入层）的增长；以人工设置的准确率/人工设置的参数量/模型不宜增长（增长后模型没有明显进步）为退出条件，使模型退出自增长；最终得到一个刚好适应对应任务的、推理效率高的成品模型。  综合研究的内容，需要解决以下四个问题：   1. 模型在什么位置时候进行深度/宽度增长 2. 模型在什么时候时候进行深度/宽度增长 3. 如何控制模型增长的停止 4. 增长的模型参数如何进行初始化   已经解决的问题：   1. 实现了随机的卷积神经网络模型深度增长，包括 rand\_grow\_mobilenet、rand\_grow\_vgg 和rand\_grow\_resnet，在各异的任务（Cifar10、Cifar100、SVHN、MNIST、ImageWoof、ImagenetTE、Mini-Imagenet、Tiny-Imagenet）上，初步实现了任务自适应，以更高的推理效率实现了超过或类似于所有固定版本的卷积神经网络的性能。 2. 研究了几种简单的新增网络结构参数初始化方式，包括原始初始化（模型自带的初始化方式）、均匀分布初始化、高斯分布初始化、全零初始化，初步以高斯初始化方式为最优。 3. 实现了基于带有结构重参数化功能的网络层，它能够自由地增加层的逻辑宽度而不会增加层的实际推理宽度。   尚未解决的问题：   1. 需要实现非随机的、可解释的、有针对性地模型结构增长位置选择。 2. 需要实现合理的模型增长停止控制方案。 3. 需要探索更加复杂的新增模型结构的初始化方法。 4. 需要在Transformer架构和自然语言处理任务上验证模型自增长的有效性。 5. 需要解决自增长模型过参数化效应不明显导致的网络难压缩问题。 | | | | | |
| 评审组意见：  本文制定出了具体的研究方法、研究路线、为论文的顺利进行打下了坚实的基础。根据目前神经网络模型体量和用户任务匹配度偏低痛点，提出通过网络自增长式训练解决，并通过结构重参数化等技术来保证网络的高性能和高效率。研究内容和进度基本和预期安排一致，论文能够达到预期的目标，论文所做的工作与开题报告内容较为符合，经评估，能在2024年4月完成论文。 | | | | | |
| 评审组成员签字： | | | | | |
| 评审组组长签字： | | | | | |

论文进展情况应包括任务要求，已经解决的和尚未解决的问题以及完成情况。评审组意见应包括对所完成内容的评价和给出是否能在明年4月完成论文。