|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **分类号** |  | **密级** |  | |
|  |  |  |  |  | |
|  | | | | | |
| **重庆邮电大学研究生学位论文**  **中期考核报告** | | | | | |
|  | | | | | |
|  | **中文题目** | **基于深度编解码器的语音识别** | | |  |
|  |  | | |
| **英文题目** | **Speech Recognition Based on Deep** | | |
|  | **Encoder and Decoder** | | |
|  |  | | |
| **学 号** | **S190301050** | | |
| **姓 名** | **程家伟** | | |
| **学位层次** | **硕士研究生** | | |
| **学位类别** | **工学硕士** | | |
| **学科专业** | **控制科学与工程** | | |
| **研究方向** | **深度学习理论及其应用** | | |
| **指导教师** | **李鹏华 教授** | | |
| **完成日期** | **2022年1月8日** | | |
|  | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **一、研究生简况** | | | | | | | | | | | | | | | | |
| 姓名 | 程家伟 | | | 性别 | 男 | | | 年龄 | | 25 | | 专业类型 | | 学术型 | | |
| 导师  /副导师或联系人 | | | | 李鹏华 | | | | 所在团队名称  团队负责人 | | | | 汽车电子与嵌入式系统理论、技术及应用重庆高校创新团队（冯辉宗） | | | | |
| 外出实习单位 | | | | 无 | | | | 外出实习时间 | | | | 无 | | | | |
| 开题时间 | | | | 2021.1.17 | | | | 开题状态 | | | | 正常 | | | | |
| 学籍异动情况 | | | | 无 | | | | | | | | | | | | |
| **二、学位论文工作进展自查** | | | | | | | | | | | | | | | | |
| **分项** | **开题报告内容** | | | | | | | | **中期完成情况** | | | | | | | |
| 论文  题目 | 基于深度编解码器的语音识别 | | | | | | | | 基于深度编解码器的语音识别 | | | | | | | |
| 研究  目标 | (1)探索基于注意力机制的语音特征序列局部上下文信息提取方法。改进单一缩放点积注意力下的局部上下文信息提取能力；兼顾模型表达能力的鲁棒性和计算效率的高效性；  (2)建立基于残差分层分组线性变换的轻量级解码器结构。给出残差分层分组线性变换的具体形式；建立其与注意力机制的拓扑关系；  (3)探索特征序列局部上下文信息和全局上下文信息的融合方法。提出多注意力机制下的多层次信息融合方法；构建多注意力编码器结构。 | | | | | | | | (1)提出了一种适用于语音特征提取的自适应掩膜局部注意力机制，进一步增强单一缩放点积注意力机制作用下的局部上下文信息提取能力；  (2)提出了基于残差分层分组线性变换的前馈网络，并将其嵌入解码器网络，进而大幅降低网络的参数量；  (3)提出了耦合全局-局部注意力机制的编码器网络结构，有效提取特征序列局部上下文信息和全局上下文信息；  (4)通过上述算法构建的网络在AISHELL-1中文普通话数据集上达到6.17%的平均字错误率，19.9M的参数量。 | | | | | | | |
| 研究  内容 | **(1)基于自适应掩膜的局部注意力机制**  针对局部上下文信息难以捕获的问题，研究基于多头注意力机制的信息重要性筛选方法，通过引入可学习参数的自适应掩膜，学习各注意力头中信息的关联范围；针对自注意力机制计算复杂度高的问题，通过可学习的线性变换表征当前目标向量和上下文向量之间的非线性联系，直接获取当前时刻各个上下文的权重，实现局部信息的高效获取。  **(2)基于全局-局部注意力机制的语音序列编码器**  针对全局和局部上下文信息难以有效融合的问题，研究基于自注意力机制和局部注意力机制的编码器结构，通过堆叠的方式融合两种注意力机制，实现全局信息和局部信息的融合；针对语音帧位置信息的表示问题，研究基于位置编码的自注意力机制，通过引入相对位置编码，实现语音帧位置信息的动态表征。  **(3)基于残差分层分组线性变换的解码器**  针对编码器-解码器模型体积庞大的问题，研究基于残差分层分组线性变换的解码器结构。通过使用低维度隐层表示和残差分层分组线性变换，耦合词嵌入(EmbeddingLayer)以及前馈网络(Feedforward Network)模块，实现端到端模型可学习参数量的降低和计算效率的提高。 | | | | | | | | 已完成研究内容(1)的算法设计和相关实验  设计了一种基于自适应掩膜的局部注意力机制：通过向局部注意力机制中引入参数化的掩膜函数，使得局部注意力机制具有动态学习各个注意力头部作用范围的能力，从而实现语音信号局部信息的高效捕获，进一步降低语音识别识别的错误率。与SpeechTransformer、HA-Transformer等仅使用自注意力机制的模型相比，本文提出的基于该注意力机制的编解码网络拥有更低的平均字错误率，在AISHELL-1中文普通话数据集的测试集上为5.65%。  已完成研究内容(2)算法设计和相关实验  通过研究全局注意力机制和局部注意力机制在不同的拓扑结构下的识别效果，提出了效果最优的“全局—局部”级联结构的融合注意力机制；提出了基于该融合注意力机制的编码器网络结构。相较于并行、嵌入等结构，提出的级联结构在AISHELL-1中文数据集和TED-LIUM2英文数据集上的字(词)错误率分别最多下降0.14和0.23。  已完成研究内容(3)的算法设计和部分实验  提出了基于残差分层分组线性变换的前馈子网络，该子网络内部采用稀疏连接，同组神经元共享同一权值矩阵，在实现较低计算量的同时大幅降低网络参数量；在研究内容(1)和研究内容(2)的基础上，设计了基于该前馈子网络的解码器网络。相较于对比方法中识别率最高的HA-Transformer，本文提出的模型参数量下降了49.6%。 | | | | | | 完成百分比  1.100%  2.100%  3.80% | |
| 创新 | (1)针对现有注意力机制无法有效提取局部上下文信息的问题，使用自适应掩膜机制对非点积注意力作用范围进行限制，使作用范围被动态地限制在某个小范围内，完成局部信息的准确、高效提取；  (2)针对全局注意力机制和局部注意力机制的融合问题，使用堆叠式结构编码器，在普通Transformer编码器自注意力模块前放置局部特征提取器，实现全局上下文信息和局部上下文信息的融合；  (3)针对普通Transformer解码器参数量多、计算量大的问题，使用基于残差分层分组变换的“梨形”单元耦合各层，大幅减少模型的参数量。 | | | | | | | | (1)针对现有注意力机制无法有效提取局部上下文信息的问题，提出了基于自适应掩膜的局部注意力机制，自适应地学习注意力作用范围，实现局部信息的准确、高效提取；  (2) 针对全局注意力机制和局部注意力机制的融合问题，提出了“全局-局部”堆叠式结构的编码器网络，实现全局上下文信息和局部上下文信息的高效融合；  (3)针对普通Transformer解码器参数量多、计算量大的问题，设计了基于残差分层分组变换的“梨形”前馈子网络，使用该子网络的编解码器网络参数量大幅降低。 | | | | | | | |
| 学术  指标 | (1)平均字错误率不高于10%；  (2)模型参数量不高于30M。 | | | | | | | | (1)平均字错误率6.17%；  (2)模型参数量19.9M。 | | | | | | | |
| 成果  指标 | 公开专利一篇，发表SCI论文一篇。 | | | | | | | | 公开专利3件，已投SCI会议论文1篇。 | | | | | | | |
| 其他  情况 | 实际采用的研究方法：  本文首先设计基于自适应掩膜和分组线性变换的编解码神经网络，然后对该网络进行训练和测试。训练部分，首先通过频谱增强技术对语音信号频谱特征(Filter Bank, Fbank)进行增强，再用堆叠的编码器(Encoder)实现高层次特征的提取，最后让标注文本的嵌入表示与编码器提取的特征通过解码器进行充分交互，得到输出序列，进而与标注文本进行比较产生损失，再通过反向传播更新网络权值。预测部分，在神经网络权值确定后，对给定的语音信号特征序列进行归一化，再送入编码器-解码器网络进行预测，之后引入额外的语言模型对输出结果进行重打分(rescore)，并使用波束搜索(Beam Search)找到输出序列的最优解。 | | | | | | | | | | | | | | | |
| 实际执行的进度安排：  2020.12~2021.01  阅读课题相关文献，整理相关理论依据并与导师交流。完成文献整理工作，撰写课题相关综述。  2021.02~2021.05  收集实验所需数据，完成部分数据收集工作。  2021.06~2021.09  设计语音识别算法模型，并进行建模，完成算法设计。  2021.10~2021.12  训练算法模型，并进行模型评估与优化。完成模型训练与优化。  2022.01~2022.03  完成毕业论文撰写。  2022.03~2022.04  论文定稿，准备答辩。 | | | | | | | | | | | | | | | |
| 中期考核时提交的材料清单（含系统演示文档、发表成果等）：  系统演示文档一份  论文一篇  专利两篇 | | | | | | | | | | | | | | | |
| 工作  小结 | 是否按开题报告执行 | | | | | | 是 | | | | | | | | | |
| 对变更开题内容的说明 | | | | | | 无 | | | | | | | | | |
| 工作进展情况说明 | | | | | | 按计划进行 | | | | | | | | | |
| 目前存在的问题和举措 | | | | | | 暂无 | | | | | | | | | |
| **三、毕业条件自查** | | | | | | | | | | | | | | | | |
| 课程  学分  情况 | | 学位课 | | 19 | | 是否完成规定学分？如无，说明原因及补救措施：  是 | | | | | | | | | | |
| 非学位课 | | 14 | |
| 成果  情况 | | 排名 | 类型  (论文、专利、获奖等) | 内容  (按文献格式) | | | | | | | 级别 | | 状态  (发表/录用/已投/计划中) | 是否论文工作相关 | | 可否获得毕业资格 |
| 2 | 论文 | adaptive mask based attention mechanism for mandarin speech recognition | | | | | | | SCI会议 | | 已投 | 是 | | 是 |
| 2 | 专利 | 李鹏华,程家伟,刘行谋,张亚鹏,俞方舟,陈旭赢,乐磊,张恩浪,董江林. 基于自适应掩膜和分组线性变换的轻量级语音识别方法: 中国,CN113113000A[P]. 2021-07-13. | | | | | | |  | | 已公开 | 是 | | 是 |
| 2 | 专利 | 李鹏华,程家伟,柴毅,程安宇,胡向东,侯杰,朱智勤,张亚鹏,董江林. 基于蒙特卡洛方法的锂电池异常工况数据自组增强方法: 中国, CN111767672A[P].2020-10-13 | | | | | | |  | | 已公开 | 否 | | 是 |
| 7 | 专利 | 李鹏华,张子健,王平,熊庆宇,邵子璇,侯杰,程家伟. 基于改进卷积-长短时记忆神经网络的锂电池容量估算方法: 中国, CN111220921A[P].2020-06-02 | | | | | | |  | | 已公开 | 否 | | 否 |
| 其他  情况 | | 如有其他和毕业相关的情况，在此说明：  无 | | | | | | | | | | | | | | |
| **四、学位论文写作计划** | | | | | | | | | | | | | | | | |
| 列出学位论文的目录计划，一般应给出全部章节的一级和二级标题（此部分对梳理学位论文工作并按期完成撰写很重要，务必在导师指导下撰写并提交）。  摘要  Abstract  目 录  第一章 绪论  1.1 研究背景及意义  1.2 国内外研究现状及存在问题  1.3 论文组织结构  第二章 端到端语音识别基础理论  2.1 语音识别系统的基本组成结构和工作流程  2.2 深度神经网络基础  2.3 基于连接时序分类的语音识别  2.4 基于编解码器模型的语音识别  2.5 本章小结  第三章 基于深度编解码器的语音识别  3.1 引言  3.2 自适应掩膜局部注意力机制  3.3 基于改进编解码器模型的语音识别  3.4 实验结果及分析  3.5 本章小结  第四章 深度编解码器模型的轻量化  4.1 引言  4.2 基于残差分组线性变换的轻量级编解码模型  4.3 实验结果及分析  4.4 本章小结  第五章 总结与展望  5.1 总结  5.2 展望  参考文献  致谢  在学期间的研究成果及发表的学术论文 | | | | | | | | | | | | | | | | |

注：以上内容根据情况添页。

