

刘 前

(年龄) 24 岁

(手机) 178-8883-3508

(求职方向) 语音算法工程师

(邮箱) liuqian.thu@foxmail.com

(GitHub) www.github.com/charlesliucn



教育经历

- | | | |
|--------------|------|-------------------|
| • 清华大学 电子工程系 | 工程硕士 | 2018.08 – 2020.07 |
| • 清华大学 电子工程系 | 工学学士 | 2014.08 – 2018.07 |
| • 清华大学 统计学中心 | 辅修学位 | 2016.08 – 2018.07 |

个人技能

- 编程语言: Python, C/C++, Shell, R, Matlab
- 使用工具: Kaldi, SriLM, KenLM, Pytorch-Kaldi, Pytorch, Keras, ESPnet
- 英语能力: 全国大学英语: 四级 (604), 六级 (556)

荣誉奖项

- | | |
|--------------------------|---------|
| • 清华大学媒体大数据认知计算研究中心学术新秀奖 | 2019.11 |
| • 国家励志奖学金 | 2016.10 |
| • 清华大学学业进步优秀奖学金 | 2015.10 |
| • 清华大学公益活动优秀奖学金 | 2015.10 |
| • 清华大学新生二等奖学金 | 2014.09 |

竞赛经历

- | | | | |
|--------------------------------------|-------|---------|---------|
| • 美国 NIST 语音技术评测, 低资源语言关键词检测赛道 | 国际第二名 | 低资源语音识别 | 2019.06 |
| • 美国 NIST 语音技术评测, 低资源语言的语音识别赛道 | 国际第三名 | 关键词检测 | 2019.06 |
| • MagicSpeechNet 家庭场景中文语音识别挑战赛, 目前排名 | 8/82 | 多通道语音识别 | 进行中 |

学术论文

- Qian Liu, Wei-Qiang Zhang, Jia Liu, Yao Liu. *A General Procedure for Improving Language Models in Low-resource Speech Recognition*, International Conference on Asian Language Processing (IALP), 2019. **LanMIT** toolkit released.
- Guan-Bo Wang, Qian Liu, Zeyu Zhao, Zhiqiang Lv, Wei-Qiang Zhang, *The THUEE Systems for the NIST Open Speech Analytic Technologies 2019 Evaluation*, NIST OpenSAT 2019 Workshop.
- Qian Liu, Wei-Qiang Zhang, Jia Liu. *Speaker Adaptation of DNN Acoustic Models using X-vectors*, submitted to Interspeech 2020.

实习经历

- 天津华控智达科技有限公司 语音识别算法实习生 2019.06 – 2020.01
 - 声学模型增量学习: 对于新语音数据, 利用原始声学模型对齐 (长音频需切分) 后, 在原模型基础上以较小学习率训练。采取迁移学习的思路, 保持原模型各层具有较小学习率, 着重对增加新神经网络层训练。
 - 解码图构建流程优化: 对 HCLG 解码图的构图方法加以更改, 采用 HCL+G 的合成方法, 以应对语言模型更新频繁的问题, 避免重复构建 HCL 部分的 WFST。
 - 语音识别后处理研究: 对语音识别系统的错误情况进行诊断分析, 添加文本纠错模块, 并使用序列标注方法恢复标点符号, 基于 OpenGRM 工具设计规则进行文本逆正则化。

- **北京华控智加科技有限公司 语音识别算法实习生** 2018.09 – 2019.05
 - **语言模型文本清洗**: 针对 400G 量级的文本数据, 使用正则表达式和数据流处理命令初步清洗数据, 并基于相对交叉熵过滤低相关度的文本, 采用多线程进行文本正则化与分词。
 - **训练集长语音切分**: 对于长音频, 对比使用两种切分方法: (1) 对齐法: 将长音频与标注做强制对齐, 在静音处进行切分; (2) 解码法: 将解码结果与标注数据匹配对应, 根据预设的时长阈值进行切分。
 - **中文发音字典构建**: 采用错误率低的 g2pC 工具获取拼音, 基于音素与拼音映射关系, 生成发音词典。
 - **语音识别系统训练**: 三万小时级声学模型的训练与优化经验。针对 400G 的通用语料, 通用语言模型训练采用并行化统计词频, 降低服务器内存压力。日语、闽南语等多个小语种的声学模型训练与优化。

项目经历

- **家庭场景多通道中文语音识别系统 系统** 2020.01 – 2020.03
 - 项目简介: 个人参加 MagicSpeechNet 语音识别竞赛, 对于多通道的家庭场景语音, 采用去混响、多通道语音融合方法及数据增强策略, 构建中英文发音词典及语言模型, 采用重打分提升识别效果。
 - 个人工作: 多通道语音的前端处理: 先用 WPE 方法进行多通道混响去除, 再使用 BeamformIt 波束形成得到单通道语音。从非语音区段提取噪声信息进行语音增强, 并使用音量扰动、速度扰动进行数据扩充。中文发音词典基于开源的 DaCiDian, 并建立中英文音素映射关系构建英文常用词发音词典。声学模型使用 TDNN-F 模型, 语言模型使用最大熵语言模型, 并用双层 Bi-LSTM 语言模型重打分。
- **语音识别后处理系统研究与设计 后处理** 2019.09 – 2020.02
 - 项目简介: 从识别结果的可读性、准确性、规范化和实时性四个角度, 研究设计语音识别后处理系统。
 - 个人工作: 基于 LSTM-CRF 序列标注模型, 添加解码的语音停顿信息, 对识别文本进行句子分割并恢复标点。使用 OpenGRM 语法工具, 设计中文文本逆正则化转换规则。对于文本纠错, 使用 N 元语言模型和 LSTM 语言模型打分后, 基于异常点检测检出错误, 并采用候选词替换和打分策略纠正错误。
- **基于 x-vector 的说话人适应训练 声学模型** 2019.08 – 2019.10
 - 项目简介: 将说话人识别的 x-vector 用于声学模型的说话人适应, 得到比用 i-vector 更优的识别结果。
 - 个人工作: 使用 TED-LIUM 语料, 根据说话人标注训练 TDNN 模型, 提取 x-vector 说话人特征。在语音识别场景下对 x-vector 提取进行优化: 训练数据筛选; 细分说话人标签以增加说话人丰富性; 去除 LDA 降维模块, 直接从网络提取所需维度特征。将 x-vector 与 MFCC 拼接进行说话人适应训练。
- **低资源条件下关键词检测评测 关键词检测** 2019.03 – 2019.06
 - 项目简介: 基于 NIST 的 OpenSAT 2019 公开评测 (语音识别和关键词检测任务), 提升低资源 (普什图语) 语音识别和关键词检测的性能指标, 重点采取多种策略提升集外关键词的检测效果。
 - 个人工作: 将前沿的跳帧 TDNN-F 声学模型应用于关键词检测任务。对于集外关键词, 一方面利用集外数据筛选高频词加入词表, 另外利用 morfessor 基于词素语言模型生成潜在新词, 降低模型的集外词比例。采用代理词策略, 提升集外词的检测性能。通过 ROVER 算法融合多个系统, 得到最优结果。
- **电信电话场景商用语音识别系统 系统** 2018.05 – 2019.01
 - 项目简介: 搭建电信电话场景商用语音识别系统, 训练使用三万小时级语音数据及 400G 通用语料库。
 - 个人工作: 采用 SRILM 工具并行化统计词频, 训练 N-gram 通用语言模型。针对应用场景, 对发音词典扩充人名和地名实体词, 并提高人名、地名等实体词的各阶频次。通用语言模型与场景语言模型进行裁剪与插值提高鲁棒性。三万小时语音为分批数据, 采用增量学习方法快速训练和更新通用声学模型。
- **低资源条件语言模型建模优化 语言模型** 2017.11 – 2018.06
 - 项目简介: 基于 NIST 的 OpenSAT 2017 评测的语音识别任务, 采取 g2p 方法扩充发音词典, 并根据文本相似度进行数据筛选, 将预训练词向量用于 LSTM 语言模型, 提升低资源场景的语言建模能力。
 - 个人工作: 对网络文本进行预处理, 用高频词扩充词表, 降低集外词比例; 使用 g2p 方法生成集外词发音词典。筛选出与训练数据语义相近的文本作为扩充数据, 方法包括: 困惑度排序、交叉熵距离排序、TF-IDF 文本相似度、基于 doc2vec 的文本相似度等。对比 Skip-Gram、CBOW、Glove 三种词向量预训练方法, 用于 LSTM 语言模型的权重初始化。优化后的语言模型帮助词错误率有 3.9% 的绝对降低。