ADILJAN ABUDUNIYAZ

Huilongguan South Rd., Changping Dist, Beijing, China
Phone:(+86)18699136853 | **Email**:adiljan1994@foxmail.com | **Homepage** https://adiladam.github.io/my_blog/About/

EDUCATION

Xinjiang University, Urumqi, China

- M.E. in Information and Communication Engineering — Sep 2018 – Jun 2021
  - Thesis: End-to-End Speech Recognition Research For Low-Resource Language
  - Adviser: **Prof. Askar Hamdulla**
- B.E. in Electronic Information Engineering — Sep 2014 – Jun 2018
  - Thesis: A research for Constructing Uyghur-Chinese Spoken Parallel Corpus

Languages Institute of Xinjiang University, Urumqi, China

- Pre-Sessional Mandarin program — Sep 2013 – Jun 2014

RESEARCH EXPERIENCE

Intelligent Information Processing Laboratory

- project: Multi-Ethnic Languages and Speech Recognition Technologies — Sep 2018 – Jun 2021
  - Supervisors: **Prof. Mijit Ablimit**, and Askar Hamdulla
  - Research topic: End-to-End speech recognition for low resource language
  - Research topic: DNN-HMM and RNN based speech recognition model
- project: Memory-augmented Chinese-Uyghur Neural Machine Translation — Sep 2017- Jun 2018
  - Supervisor: Prof.Askar Hamdulla
  - Research topic: Parallel Corpus, Low resource Language, Text Processing.

PUBLICATIONS

JOURNALS

[1] Adiljan Abuduniyaz, Mijit Ablimit, and Askar Hamdulla, "Uyghur Speech Recognition Based on DNN-HMM and RNN ," *Modern Electronics Technique*, vol. 5, pp. 90–94, Oct 2021.

CONFERENCES

[1] Adiljan Abuduniyaz, Mijit Ablimit, and Askar Hamdulla, "The Acoustical and Language Modeling Issues on Uyghur Speech Recognition," *2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Xi'an,China pp.366-369, Oct 2020.

WORK EXPERIENCE

SinoVoice Technology, Beijing, China — Mar 2024 –

- Research Engineer, AI Research Department
  - **ASR model adaptation for medical field usage**
    Mandarin is full of numerous homophonic words, as a result, our current model got worst performance on the medical field recognition task. In order to improve the model's robustness for medical data while maintaining it's capability for common recognition, conducting two phase optimization strategy.One is increasing high quality medical field data while finetuning. the second is doing model average. by the way, our model obtained remarkable improvement on both common and medical test set.
  - **Streaming batch inference and export onnx model**
    ASR model's inference time, real time factor(RTF), is one of the critical indicators. there have been various ways to reach out. In this stage, I focus on the streaming batch inference and onnx converting. Fulfilled a dynamic batch streaming conformer model.
  - **punctuate restoration modeling for ASR results**
    In general, the ASR system has no punctuation restoration ability, as a result the text which the ASR system produced encounters readability degradation. To resolve the above issue,cascading a punctuation restoring model is a primary method. I've tried in terms of models for this task, such as BERT, ct-transformer.Among these models, ct-transformer showed the promising performance both streaming and non-streaming scenario.

- **Multi CPUs ASR model training on Kylin Linux**
  Not all of AI models might be trained on GPUs, such a place where high data security demanded, the CPUs is the only way one must go through.Herein, I utilize Wenet as the training tool, and implemented multiple CPUs training with pytorch data parallel method and packaged it as Docker image so that client can easly training their own models on Kylin Linux OS.
- **Mixture of Expert with Conformer Model**
  Mixture of Expert with Conformer Model has shown competitive performance on large speech data with its better real time factor and promising WER compare to 1B conformer model[**1**,**2**]. Therefore, in this stage, I've been training MoE+conformer and 1B conformer model on 120k hours mixed English and Chinese speech data.The experimental results show that the conformer combined with MoE model has proven competitive performance on various test sets with favorable inference time compare to 1-B conformer model.
- **Minority language ASR Model Optimizing**
  Align with client demand, the previous version of Uyghur ASR model has need to be upgraded and further enhanced its capacity, its actual word error rate(WER) is required to be decreased from 23% to 15%. In this section, I have completed two types of further optimization for the above demand.Firstly, the text data mapped into Latin from Arabic, as well as acoustic units extended from 512 BPEs to 2000 BPEs. Secondly, increased 100 hours high quality training data. In doing so, model has obtained 2.5%-3.3% actual WER reduction in test sets.
- **Multi GPUs Multi Machine ASR model training**
  To further facilitate of the performance of ASR system, some novel architecture must be taken into account as well as large amount of data. in order to meet above two demand, I chose branchformer as the ASR model, deepspeed as the distributed training tool, using approximately 120k hours mixed Chinese and English speech data and multi GPUs multi machines(8 machine and each of it has 8*3080 GPUs), trained a ASR model, then evaluate the model on various test sets.
- **Semi-Supervised Multilingual ASR model**
  Developed multilingual speech recognition system based on WavLM Architecture, multilingual dataset includes 25 languages,which covers most common languages and some accents, such as English, Chinese, Arabic, Japanese, Cantonese,etc.My job in this part was data processing, downstream fine-tuning, model evaluation and Docker Packaging.Data processing step is mainly operating text signal and it can be further separated data selecting, determining acoustic units, generating BPE model and training language model. Downstream task takes place by WavLM-CTC ASR on each of languages which approximately had 200 400 hours of speech data.These fine-tuned models, then, evaluated on 8k and 16k test sets respectively. Lastly, to meet the client's demand the experimental environment packaged into docker, online decoding capacity implemented with VAD, both wave files and microphone streams also can be the inputs of the ASR system.

JunLin Technology, Suzhou, China Feb 2023 – Mar 2024

- Machine Learning Engineer, Research & Development Division
  - **Voice Conversion**
    Constructing training data set for voice conversion model. Specifically, obtained target waves from online at first step. Then, detached silence from wave files, cut it into small length wave files according to the time stamps. Finally, generated good confidence wave files, which for training voice conversion model, passing the small length waves through speech enhancement model. Studying some mainstream voice conversion model, such as GPT-SoVITS.
  - **Keyword Spotting Model**
    Conducted temporal convolution network(TCN) based keyword spotting model with Max-pooling, Cross-entropy(CE) and CTC loss as target function. Herein, Max-pooling and CE are fitting positive and negative labels while CTC align character ground truth. Then in order to adapting custom keyword spotting, take the CTC-based model into account, fine-tuned this model on small amount of custom data, which includes two keywords i.e. 'nihaolele', 'xiaoyixiaoyi'. Generate Micro service of KWS using google gRPC.
  - **Speech Enhancement Model**
    Investigate and survey the most recent advancements in speech enhancement realm and speech-related research through rigorous literature reviews. Implement two speech enhancement models base on CNN-LSTM and FRCRN respectively, in streaming and non-streaming style. then developed its python API to support ASR system or to building data set.
  - **Voice Activity Detection Model Based on Deep Learning**
    During this stage, inquired into the latest improvements in facet of voice activity detection and comprehensively reviewing related research articles. Afterwards, on the basis of Silero and Pyannote, developed two VAD systems, which are able to processing speech in streaming and non-streaming way, offered to help with ASR system working efficiently.
  - **End-to-End Speech Recognition Model and it's Serialization**
    Studied conformer based End-to-End speech recognition model. With a primary focus on Wenet toolkit, research and develop an end-to-end automatic speech recognition system with multiple decoding methods. Then, in order to accelerate decoding speed, and to decrease calculating costs, converted the PyTorch model into TorchScript and ONNX formats.

SpeakIn Technology, Shanghai, China Dec 2021 – Dec 2022

- Data Scientist, Research Academy

- **Study Unsupervised Feature Extracting Methods**
  With the aim of broadening understanding of the most recent advancements in the field of speech recognition, reviewed research papers related to unsupervised feature extraction, contrastive learning, and semi-supervised learning, etc. Conducted some experiments on Fairseq toolkit to validate unsupervised learning methods.
- **Punctuation Restoration Model Training and Optimizing**
  Developed semi-supervised Transformer-based model for punctuation restoration in end-to-end Uyghur language speech recognition system. Initially, constructed a multi-class model, with period, comma, exclamation point, and question mark designated as the target classes. The primary model structure employed for this purpose was the Transformer-encoder,sub-word was the unit. Unfortunately, this model showed weak competitive results on test data set. Subsequently, to increasing performance of the model, substituted modeling unit into byte level sub-word, additionally trained a masked language model. As the downstream model, build a classifier model and combine with masked language model, then fine-tuned. As a consequence, fine-tuned model has showed most reliable results on test set.
- **End-to-End ASR Model Training and Optimizing**
  Made a profound understanding of end-to-end speech recognition, which includes models based on transformer and conformer architectures, as well as various decoding methods such as greedy search, beam search and attention re-scoring. trained several end-to-end ASR model using substantial open-source Uygur speech recognition dataset comprising over 150 hours of data. Conducted resolution studies to assess the effectiveness of various modeling units within the models. Statistical experiments provided a trustworthiness to determining which modeling units are more suitable for Uyghur language speech recognition model. Furthermore, I have been exploring the optimization of neural networks, quantization techniques, and serialization methods in my learning process.
- **Acoustical and Language Data Processing**
  To assist the team in establishing the Uyghur speech and text dataset required for speech recognition, including specifying data sources, inspecting data quality, data cleaning, and training language model,etc.

AWARDS & HONORS

- Autonomous Region Graduate Scholarship — Jun 2021
- Outstanding Graduates of School of Information Science and Engineering — Jun 2018
- Third Prize in the Software Service Competition of the 10th Chinese College Student Computer Competition — May 2017
- The ”Excellent Award” of the 11th Challenge Cup Extracurricular Science and Technology Works Competition for College Students — May 2015
- Merit Student of Xinjiang University — 2015 - 2016
- National Encouragement scholarship — 2015 - 2016
- Merit Student of Xinjiang University — 2014 - 2015
- National Encouragement scholarship — 2014 - 2015
- Outstanding Class Cadre of the Language School of Xinjiang University — 2013 - 2014

TECHNICAL STRENGTHS

- OS: Linux, Windows
- Programming Languages: Python,Shell; C/C++
- Deep Learning Frameworks: PyTorch, Keras, Tensorflow
- Deep Neural Models : Transformers, BERT, Wav2vec, WavLM,CNNs, RNNs
- Deploy: Docker, gRPC, Torchscript, ONNX, TensorRT
- ASR tool: Kaldi,Espnet,Wenet,fairseq,FunASR
- Platform: Huggingface, Github, Kaggle, Confluence

COMMUNICATING

- Uyghur: Native
- Mandarin: Fluent
- English: proficiency

INTERESTS

- Cycling
- Swimming
- Digital photography
- Cooking