Dongwei Jiang

✉ jiangdongwei0@gmail.com ☎ 443-453-7610 Some-random 🎓 Google Scholar

With six years of industry and research experience in speech processing and self-supervised speech models, my focus is now shifting to LLM, with a particular interest in enhancing its reasoning capabilities.

EDUCATION

JOHNS HOPKINS UNIVERSITY	**M.Eng. in Electrial and Computer Engineering**	Aug 2023 - Till Now
PEKING UNIVERSITY	**B.S. in Geographic Information System**	Sep 2011 - June 2016

RESEARCH EXPERIENCE

JOHNS HOPKINS UNIVERSITY Baltimore, Sep 2023 - Till Now

Working with Professor Daniel Kashabi and Benjamin Van Durme on natural language reasoning

- Trying to address complex reasoning challenges by treating them as planning problems. Leveraging the assessment capabilities of Large Language Models (LLMs) to evaluate states and provide environment feedback. Employing planning techniques such as Monte Carlo Tree Search (MCTS) to navigate solutions within various environments. My current focus is on developing verification module that is sufficiently versatile to apply across a broad range of reasoning issues, aiming to overcome the shortcomings in existing approaches like Reflexion
- Assisted in the problem decomposition and search error analysis of open-domain NELLIE - a system that provides grounded explanations/proofs for answer in general open-domain question answering settings

THE UNIVERSITY OF EDINBURGH Remote, Dec 2022 - Till Now

Working with Professor Shay Cohen, conducted research on natural language logical reasoning with LLM + theorem prover. The paper got an average review of 3.5 and a meta-review of 4 for ARR October

- Devised a structured approach for solving natural language reasoning tasks using the theorem prover Lean, by dividing the task into formalization and proving stages
- Annotated proof data for natural language reasoning problems from ProofWriter and FOLIO. Trained a premise retriever and a retrieval-based tactic generator with and without pre-training on Lean's theorem proving data. Achieved state-of-the-art results on FOLIO and ProofWriter

DIDICHUXING TECHNOLOGY Beijing, Jan 2018 - Dec 2020

Advancing self-supervised speech recognition methods with MPC and Speech-SimCLR

- Proposed unsupervised speech recognition pre-training methods Masked Predictive Coding (MPC) [4], one of the first to apply BERT-like reconstruction loss to Speech Transformer for unsupervised pre-training. Extended MPC to streaming scenarios [2] and different downstream tasks [3]. Investigated domain mismatch between pre-training and fine-tuning
- Proposed Speech-SimCLR [1], one of the first framework that combines contrastive and reconstruction objectives for unsupervised speech representation learning. It employed various speech augmentations and achieved competitive results on multiple speech tasks such as speech recognition and speech emotion recognition
- Led the effort to refactor DiDi Chuxing's internal speech framework code base and released it on Github (project name Athena). Athena contains industrial engine for ASR/TTS/ASV/VC, various unsupervised pre-training algorithms and WFST construction/decoding pipeline with end-to-end framework. Currently, the project has 800+ stars and 200+ PRs

SELECTED PUBLICATIONS

[1] **Dongwei Jiang**, Wubo Li, Miao Cao, Wei Zou, Kun Han and Xiangang Li, Speech SIMCLR: Combining Contrastive and Reconstruction Objective for Self-supervised Speech Representation Learning, in InterSpeech 2021

[2] **Dongwei Jiang**, Wubo Li, Ruixiong Zhang, Miao Cao, Ne Luo, Yang Han, Wei Zou and Xiangang Li, A Further Study of Unsupervised Pretraining for Transformer Based Speech Recognition, in ICASSP 2021

[3] Ruixiong Zhang, Haiwei Wu, Wubo Li, **Dongwei Jiang**, Wei Zou and Xiangang Li, Transformer Based Unsupervised Pre-Training for Acoustic Representation Learning, in ICASSP 2021

[4] **Dongwei Jiang**, Xiaoning Lei, Wubo Li, Ne Luo, Yuxuan Hu, Wei Zou and Xiangang Li, Improving Transformer-based Speech Recognition Using Unsupervised Pre-training, CoRR abs/1910.09932

WORK EXPERIENCE

SHOPEE Beijing, Nov 2021 - Nov 2022

Led a team of four. Responsible for low-resource ASR and its application in products

- Low Resource ASR Framework
  - Investigated multiple techniques to optimize low-resource ASR performance. Used Youtube data crawling and filtration to accumulate weakly supervised training data for base model (+15% on base model, 100k hours of data for each language), self-training to accumulate target domain training data (+10% on target domain, 10k hours of data for each domain), RandAugment and meta learning to improve accuracy in the training stage
  - Established best practice for low resource ASR annotation. Enforced multiple rounds of cross-validation at the annotation stage to ensure the correctness of annotation result with minimum extra effort. Trained local managers to help local agents understand annotation rules and supervise the whole process
- Product Support
  - Supported video search, video quality check and video understanding for ShopeeVideo on ID market. Video search got +0.1 DCG with the addition of ASR information. Video QC got 80% recall rate on speech-related test set
  - Supported live streaming clipping and live streaming ongoing content search on ID market. Achieved 95% acceptance rate for live stream clips with regard to relevance, exquisiteness, and user experience criteria raised by the business team. Used as an extra source of supply for product introductory videos, it increases 23,000 orders each day

YUANFUDAO Beijing, Dec 2020 - Oct 2021

Led a team of six. Responsible for research and application of End-to-End ASR and Talking Face Generation

- End-to-End ASR for Education
  - Built end-to-end ASR framework for Yuanfudao from the ground up with non-streaming Speech Transformer and streaming CTC + Attention rescoring two-pass framework. Collaborated with HPC team and achieved 10x speedup on Speech Transformer with beam search optimization and Cuda-level operator reimplementation
  - Investigated and implemented relevant techniques for education applications, including code switch (for English courses), children speech recognition, and deep contextual biasing (on decoder, for proper nouns in subjects)
  - Supported quality check of Zebra Shot (children short video application) and live lesson streaming. The overall recall rate on speech test set reached 82% with under 3% review rate. The whole project saves the company 900,000 RMB from manual quality check each month
  - Supported content understanding of voice communication between teachers and parents. With the collaboration from NLP team, the recall rate of the service violation/cheating/over-commitment reached 70% with 10% review rate, while the intent classification accuracy of service refund reached 93%
- Talking Face Generation
  - Investigated well-known 2D/3D Talking Face Generation approaches like ATVG, DAVS, and VOCA. Implemented Wav2lip and experimented with different input features on Wav2lip model

DIDICHUXING TECHNOLOGY Beijing, Jan 2018 - Dec 2020

Led a team of five. Responsible for the research and training/deployment platform of ASR and TTS

- End-to-End Speech Recognition Framework
  - Implemented LAS and Speech Transformer with Attention + CTC multi-task loss. Speech Transformer achieved more than 20% relative WER reduction compared to CTC baseline while also being faster. Being one of the first to bring end-to-end ASR model online in China, Speech Transformer replaced all hybrid/CTC models for internal products
  - Implemented AED + CTC joint decoding and Transformer LM rescoring for general use case. Implemented word-level WFST with attention decoder and contextual-graph based hotword-fix capability for personalized use case
  - Investigated and implemented prevalent end-to-end streaming speech recognition solutions at the time (Neural Transducer, RNN-T, and Mocha). Neural Transducer using character as modeling units got lower WER than CTC model with similar latency. RNN-T with fast-emit got similar performance and lower latency than CTC baseline
- End-to-End Speech Synthesis Framework
  - Implemented Tacotron2 and Fastspeech end-to-end Speech Synthesis model. FastSpeech achieved 4.52 MOS score over 4.34 from previous parametric model while also being faster
  - Implemented multi-speaker TTS and GST for personalized TTS. Trained on internal 1500 people low-quality TTS dataset, the similarity score for seen/unseen speaker reached 3.42/3.03

JD.COM INC Beijing, Jan 2017 - Dec 2017

- Read through Kaldi source code. Learned the concept and implementation of hybrid ASR system. Followed early CTC work from Google and experimented with different network structures, modeling units, and subsampling rate