Rao Ma

(+44)7759-759617 | rm2114@cam.ac.uk

Research Interest

Automatic Speech Recognition, Domain/Task Adaptation, Natural Language Processing

Education

University of Cambridge Oct 2022 – Present

PhD, Information Engineering

- First-author publications at top-tier NLP/Speech conferences, e.g. NAACL, Interspeech
- Work in the Machine Intelligence Lab (MIL) under the supervision of Prof. Kate Knill
- Research Topic: Domain and Task Adaptation of Foundation Speech Models

Shanghai Jiao Tong University Sep 2018 – Mar 2021

M.S., Computer Science and Technology

- GPA: 3.86/4.00, Rank: 11/77
- Work in the Cross Media Language Intelligence Lab (X-LANCE) under the supervision of Prof. Kai Yu
- Thesis: Application of Neural Network Language Model in Speech Recognition

Nanjing University Sep 2014 – Jun 2018

B.S., Computer Science and Technology

- GPA: 4.58/5.00, Rank: 1/150
- Work in the Websoft group under the supervision of Prof. Yuzhong Qu
- Thesis: Study on Language Modeling based on Teacher-Student Learning

Industry Experience

ByteDance Apr 2021 – Aug 2022

Full-time Speech Recognition Engineer

- Developed NNLM solutions in a LAS-based ASR system for Douyin and TikTok speech recognition services
- Investigated and implemented the novel Internal Language Model Estimation and Training algorithm in a LAS-based ASR system, and proposed several effective variants to the original algorithms
- Developed an optimization pipeline for the ASR system to improve long-tail proper noun recognition in automatic caption generation for TikTok and CapCut

AISpeech Apr 2020 – Oct 2020

Speech Synthesis Engineer Intern

- Developed a large-scale PyTorch-TTS training framework supporting Tacotron 2, FastSpeech, and LSTM-TTS
- Participated in the Interspeech 2020 Accented English Speech Recognition Challenge (AESRC), and won the first and second places in two challenge tracks as a team member
- Applied a phoneme-based algorithm to select Chinese dialect corpus for recording TTS training data

Research Experience

University of Cambridge Oct 2022 – Present

MIL with Prof. Kate Knill

- **Investigate the Emergent Ability of Whisper**
  - Proposed the usage of template-based text prompt and debiasing for zero-shot audio classification
  - Whisper showed emergent audio classification ability on 8 datasets, outperforming previous methods by 9%
- **Task Adaptation of Whisper**
  - Adopted fine-tuning, LoRA tuning, and soft prompt tuning to adapt Whisper for downstream tasks
  - Introduced end-to-end approaches for disfluency removal and spoken grammatical error correction
- **ASR Error Correction with Foundation Language Models**
  - Investigated supervised ASR error correction methods with N-best T5 and zero-shot approaches with LLMs
  - Experiments showed promising results on outputs from both Transducer and AED-based ASR models

Shanghai Jiao Tong University Sep 2018 – Mar 2021

X-LANCE with Prof. Kai Yu

- **Neural Lattice Search for Automatic Speech Recognition**
  - Proposed a novel Lattice-to-Sequence model with push-forward algorithm for the ASR second-pass decoding
  - Yielded 9.7% and 7.5% relative WER reduction compared to N-best rescoring and lattice rescoring
- **Large-Scale Language Model Compression**
  - Combined product quantization, binarization and knowledge distillation training for LSTM LM compression
  - Obtained an overall compression ratio of 100 on standard datasets with little performance degradation
- **Investigation on Multi-Sense Word Embedding of NNLM**
  - Proposed an extension to NNLM called "Structured Attentional Multi-Sense Embeddings" that learned multiple fine-grained embeddings for each word in an unsupervised manner
  - Conducted qualitative and quantitative analysis to show the mitigation of the meaning conflation deficiency

Nanjing University Sep 2016 – Jun 2017

Websoft with Prof. Yuzhong Qu

- **Question Entity Discovery and Linking**
  - Optimized the speed of question entity linking algorithm in a general domain question answering bot
  - Built a general domain question entity discovery system with CRF model, and ranked 11th in CCKS-2017

Selected Publications

**Google Scholar:* https://scholar.google.com/citations?user=4jn7KMIAAAAJ

Journal Articles

[1] Qi Liu, **Rao Ma**, and Kai Yu. "Markov Decision Process and Prior Control Vector for Weak Condition Natural Language Generation". *Chinese Journal of Computers* 45 (2022), pp. 290–301.

[2] Kai Yu, **Rao Ma**, Kaiyu Shi, and Qi Liu. "Neural Network Language Model Compression With Product Quantization and Soft Binarization". *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), pp. 2438–2449.

[3] Su Zhu, Zijian Zhao, **Rao Ma**, and Kai Yu. "Prior Knowledge Driven Label Embedding for Slot Filling in Natural Language Understanding". *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), pp. 1440–1451.

Articles in Conference Proceedings

[4] **Rao Ma***, Adian Liusie*, Mark JF Gales, and Kate M Knill. "Investigating the Emergent Audio Classification Ability of ASR Foundation Models". In: *NAACL 2024*.

[5] Mengjie Qian, **Rao Ma**, Adian Liusie, Erfan Loweimi, Kate M Knill, and Mark JF Gales. "Zero-shot Audio Topic Reranking using Large Language Models". In: *arXiv preprint arXiv:2309.07606*.

[6] **Rao Ma**, Mengjie Qian, Potsawee Manakul, Mark Gales, and Kate Knill. "Can Generative Large Language Models Perform ASR Error Correction?" In: *arXiv preprint arXiv:2307.04172*.

[7] Stefano Bannò, **Rao Ma**, Mengjie Qian, Kate M Knill, and Mark JF Gales. "Towards End-to-End Spoken Grammatical Error Correction". In: *ICASSP 2024*.

[8] **Rao Ma**, Mengjie Qian, Mark JF Gales, and Kate M Knill. "Adapting an ASR Foundation Model for Spoken Language Assessment". In: *SLaTE 2023*.

[9] **Rao Ma***, Mengjie Qian*, Mark JF Gales, and Kate M Knill. "Adapting an Unadaptable ASR System". In: *Interspeech 2023*.

[10] **Rao Ma**, Mark JF Gales, Kate M Knill, and Mengjie Qian. "N-best T5: Robust ASR Error Correction using Multiple Input Hypotheses and Constrained Decoding Space". In: *Interspeech 2023*.

[11] **Rao Ma**, Xiaobo Wu, Jin Qiu, Yanan Qin, Haihua Xu, Peihao Wu, and Zejun Ma. "Internal Language Model Estimation Based Adaptive Language Model Fusion for Domain Adaptation". In: *ICASSP 2023*.

[12] Yufei Liu, **Rao Ma**, Haihua Xu, Yi He, Zejun Ma, and Weibin Zhang. "Internal Language Model Estimation Through Explicit Context Vector Learning for Attention-based Encoder-decoder ASR". In: *Interspeech 2022*.

[13] Tian Tan, Yizhou Lu, **Rao Ma**, Sen Zhu, Jiaqi Guo, and Yanmin Qian. "AISpeech-SJTU ASR System for the Accented English Speech Recognition Challenge". In: *ICASSP 2021*.

[14] Houjun Huang, Xu Xiang, Yexin Yang, **Rao Ma**, and Yanmin Qian. "AISpeech-SJTU Accent Identification System for the Accented English Speech Recognition Challenge". In: *ICASSP 2021*.

[15] **Rao Ma**, Hao Li, Qi Liu, Lu Chen, and Kai Yu. "Neural Lattice Search for Speech Recognition". In: *ICASSP 2020*.

[16] **Rao Ma**, Lesheng Jin, Qi Liu, Lu Chen, and Kai Yu. "Addressing the Polysemy Problem in Language Modeling with Attentional Multi-Sense Embeddings". In: *ICASSP 2020*.

[17] Zihan Zhao, Yuncong Liu, Lu Chen, Qi Liu, **Rao Ma**, and Kai Yu. "An Investigation on Different Underlying Quantization Schemes for Pre-trained Language Models". In: *NLPCC 2020*.

[18] Ruisheng Cao, Su Zhu, Chenyu Yang, Chen Liu, **Rao Ma**, Yanbin Zhao, Lu Chen, and Kai Yu. "Unsupervised Dual Paraphrasing for Two-stage Semantic Parsing". In: *ACL 2020*.

[19] **Rao Ma**, Qi Liu, and Kai Yu. "Highly Efficient Neural Network Language Model Compression Using Soft Binarization Training". In: *ASRU 2019*.

Services

Reviewer

- IEEE Transactions on Audio, Speech, and Language Processing 2022, 2023
- IEEE Transactions on Big Data 2023

Invited Talks

- Technical Sharing with Nvidia AI/ML Specialist Team, Jan 2023
  Topic: Emergent Audio Classification Ability of Foundation Speech Models
- ALTA Technology Seminar with Cambridge University Press & Assessment, Oct 2023
  Topic: Adapting Whisper for Spoken Language Assessment and Feedback

Honors & Awards

Challenges

- 2^{nd} place, AESRC Track 1: Accented English Speech Recognition, Interspeech 2020
- 1^{st} place, AESRC Track 2: English Accent Identification, Interspeech 2020
- Meritorious Winner, Mathematical Contest in Modeling, 2016

University Scholarships

- CUP&A Studentship (Full funding for PhD), University of Cambridge, 2022 – 2026
- KLA-Tencor Scholarship, Shanghai Jiao Tong University, 2021
- Zhenggang Scholarship, Nanjing University, 2017
- National Scholarship, Nanjing University, 2015

Awards

- Excellent New Employee, ByteDance AI Lab, 2021
- Outstanding Graduate of Shanghai Jiao Tong University, 2021
- Outstanding Graduate of Nanjing University, 2018
- Merit Student of Jiangsu Province, 2017
- Outstanding Student of Nanjing University, 2016