MUTIAN HE

mutian.he@idiap.ch | Homepage: mutiann.github.io | Google Scholar | DBLP | Github

EDUCATION

Idiap Research Institute, École Polytechnique Fédérale de Lausanne (EPFL) | PhD

Martigny, Switzerland | Advised by Phil Garner | June 2022 –

The Hong Kong University of Science and Technology | MPhil

Hong Kong SAR | Advised by Yangqiu Song | Sept 2019 – June 2022 | GPA 4.03/4.3

Beihang University (BUAA) | B.Eng. in Computer Science

Beijing, China | Sept 2015 – Jun 2019 | GPA 3.8/4.0 (top 5%)

University of Toronto | Exchange Student

Toronto, Canada | 2017 | GPA 4.0/4.0

RESEARCH EXPERIENCES

Idiap Research Institute, EPFL | PhD Student | June 2022-

Working on spoken language understanding with the goal of capturing high-level semantics from speech, possibly aided by large language models and Bayesian transfer learning approaches:

Mutian He, Philip N. Garner. The Interpreter Understands Your Meaning: End-to-end Spoken Language Understanding Aided by Speech Translation. *Findings of EMNLP 2023*

Mutian He, Philip N. Garner. Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding. *Interspeech 2023*

Speech Output Scientists, Microsoft | Intern | Jul 2018 – Aug 2019; Aug 2020 – Oct 2021

Conducted multiple research projects on end-to-end neural TTS, including:

• Building robust end-to-end neural TTS systems, based on specially-designed attention mechanisms, data augmentation, and dual-learning of ASR and TTS. Stepwise Monotonic Attention achieves significantly improved robustness:

Mutian He, Yan Deng, Lei He. Robust Sequence-to-Sequence Acoustic Modeling with Stepwise Monotonic Attention for Neural TTS. *Interspeech 2019* [Demo]

• Scalability-centered low-resource TTS, which focused on extending neural voices to amounts of languages with a unified framework and minimal per-language effort, by leveraging multilingual transfer learning and discovery of a unified phoneme inventory.

A multilingual TTS model for 40+ languages achieves adaptation to new languages with <1min data without language-specific G2P or other expertise, and its mechanism is further explored:

Mutian He, Jingzhou Yang, Lei He, Frank K. Soong. Multilingual Byte2Speech Models for Scalable Low-resource Speech Synthesis. *arXiv: 2103.03541* [Demo]

For the difficulty to handle non-phonemic scripts in fully end-to-end low-resource TTS, a neural lexicon reader model is further proposed, to leverage raw textual knowledge, which avoids building G2P pipeline for each language, and gives better controllability:

Mutian He, Jingzhou Yang, Lei He, Frank K. Soong. Neural Lexicon Reader: Reduce Pronunciation Errors in End-to-end TTS by Leveraging External Textual Knowledge. *Interspeech 2022* [Demo]

• Familiarity and hands-on experiences in other topics in deep learning, including pretrained models and generative models, and various topics on speech including ASR, voice conversion, neural vocoding, as well as speaker and prosody modelling.

HKUST KnowComp Group | MPhil Student

Working on construction and processing of large-scale commonsense knowledge graph with

NLP techniques, thesis focused on commonsense reasoning by concept abstraction and instantiation and crowdsourced dataset on conceptualization, with papers:

Mutian He, Tianqing Fang, Weiqi Wang, Yangqiu Song. Acquiring and Modelling Abstract Commonsense Knowledge via Conceptualization. *arXiv: 2206.01532*

Mutian He, Yangqiu Song, Kun Xu, Dong Yu. On the Role of Conceptualization in Commonsense Knowledge Graph Construction. *arXiv: 2003.03239*

Xin Liu, Haojie Pan, **Mutian He**, Yangqiu Song, Xin Jiang. Neural Subgraph Isomorphism Counting. *KDD-2020*

Yu He, Jianxin Li, Yangqiu Song, **Mutian He**, Hao Peng. Time-evolving Text Data Classification with Deep Neural Networks. *IJCAI-ECAI-2018*

SKILLS

Coding: Proficiency in various programming languages from the high-level design to the hardware, with experience in competitive programming

Machine Learning: Theoretical knowledge and rich practical experiences on deep learning from data collection to implementation and training know-hows

Natural Language and Speech Processing: Strong background in NLP and Speech, with multiple projects published based on individual ideas and codes

Research: Understanding in the research pipeline, with good paper-writing skills

Language: Mandarin (Native), English (Advanced; TOEFL 111, GRE 334+4.0)

RELEVANT COURSEWORK

Parallel Programming (HKUST, A+): Topics on parallel programming, including competitive projects using pthread, MPI, and CUDA

Optimizations for Machine Learning (HKUST, A-): Theoretical discussions on convex optimization methods under the context of statistical learning

Deep Learning, Towards Deeper Understanding (HKUST, A): Discussions on recent findings on deep learning theories, including optimizations, generalizations, overparameterization, neural tangent kernels, and implicit regularization

Computer Organization (BUAA, 90/100): Discussions on computer architectures, including implementation of a full-fledged 5-stage MIPS CPU from scratch in Verilog

Various courses on diverse aspects of computer sciences, including cryptography and security, computer networks, OS, databases, compilers, computer vision, etc.

HONORS & AWARDS

Beihang Award of Academic Excellence, 1st Prize, 2015-2017

Beihang Award of Academic Contests, 1st Prize, 2015-2017

Beihang Outstanding Student of the Year, 2015-2017

Ministry of Education - CSC Outstanding Undergraduate Exchange Scholarship, 2017

HKUST Postgraduate Studentship

TEACHING

TA of COMP4221 Intro to Natural Language Processing, HKUST, Spring 2020

TA of M06 Intro to Speech Processing, Idiap, Fall 2022