

MAOHAO SHEN

550, Memorial Drive, Cambridge, 02139

217-693-9251 | maohao@mit.edu | maohaos2.github.io/Maohao

EDUCATION

University of Illinois at Urbana-Champaign

Aug 2017-May 2021

Major: Computer Engineering (B.S.), Minor: Mathematics

Champaign, IL

- GPA: 4.0/4.0

- Graduated with the Highest Honors & Bronze Tablet

Massachusetts Institute of Technology

Sept 2021-May 2023

M.S. in Electrical Engineering and Computer Science

Cambridge, MA

- GPA: 5.0/5.0

Massachusetts Institute of Technology

Sept 2023-Present

Ph.D. in Electrical Engineering and Computer Science, Minor: Mathematics

Cambridge, MA

- GPA: 5.0/5.0

RESEARCH INTERESTS AND OBJECTIVE

I focus on the development of theory-driven methodologies for addressing machine learning challenges, with a primary emphasis on enhancing trustworthiness in machine learning systems. This includes various aspects, including but not limited to uncertainty quantification and fairness. Furthermore, my recent endeavors have involved improving the interpretability, reliability, and efficiency of Large Language Models (LLMs).

WORK EXPERIENCE

IBM Research & MIT-IBM Watson AI Lab

June 2023-Sept 2023

Research Intern

Cambridge, MA

- We propose a novel method named *Thermometer* for calibrating large language models that is computationally efficient, preserves accuracy, and generalizes well to new tasks.
- Results show that our developed method generalizes well across different multiple-choice question-answering tasks and various large-scale language models.
- We further extend the proposed *Thermometer* method to the calibration of free-form generation tasks.
- The developed method will be applied to IBM internal LLMs for practical application.

IBM Research & MIT-IBM Watson AI Lab

June 2022-Sept 2022

Research Intern

Cambridge, MA

- Implement various prompt tuning methods for the large language models under the practical black-box setting.
- Propose novel and efficient algorithms to optimize prompts without gradient backpropagation or accessing the internal likelihood of LLMs.
- Towards trustworthy prompt tuning by injecting the uncertainty quantification capability to prompt tuning using Bayesian techniques.
- Demonstrate the empirical solid performance of our proposed methods in terms of prediction accuracy and the capability to solve downstream uncertainty quantification tasks such as text OOD detection.

ACADEMIC RESEARCH

Massachusetts Institute of Technology

Sept 2023-Present

Graduate Research Assistant -PhD

Cambridge, MA

- Improve Evidential Deep Learning:

- Provide a unified view of modern evidential deep learning (EDL) methods for uncertainty quantification.
- Discover the shortcomings of existing EDL methods in the learned epistemic uncertainty through theoretical analysis.
- Develop a new EDL model based on a mixture of Dirichlet distributions with Bayesian learning.

- Trustworthy Learning for LLMs:

- Explore various aspects of improving the reliability of LLMs.
- Propose novel methods for calibrating LLMs and enhancing their uncertainty quantification performance.
- Study the problem for efficient inference for LLMs leveraging uncertainty estimation techniques.

- Source-free Domain Adaptation:

- Motivated by privacy concerns, we tackle a domain adaptation problem without source domain data.
- Conduct a theoretical investigation and develop an information theoretical generalization error bound to reveal an inherent bias and variance trade-off.
- Propose a theory-driven novel methodology that outperforms the existing methods.

- Post-hoc Uncertainty Quantification:

- Tackle the uncertainty quantification problem under the post-hoc setting.
- Propose a novel approach that trains a Bayesian meta-model to help improve the uncertainty quantification performance of arbitrary pre-trained models in a post-hoc manner.
- Results show that the proposed method achieves SOTA performance on downstream OOD detection and selective classification tasks.

- Active Learning for Brain Tumor Segmentation:

- Propose a novel active learning framework tailored for the 3D brain image segmentation problem.
- Formulate the batch active learning problem as a constrained combinatorial optimization problem and develop an efficient optimization algorithm.

- Bath Active Learning:

- Propose a novel batch active learning algorithm from the perspective of sparse approximation.
- The proposed algorithm explicitly balances the uncertainty and diversity in active learning applications.
- The proposed method can adapt to various settings, including Bayesian and Non-Bayesian neural networks.

- Dimensionality Reduction for Image Processing:

- Study the problem of manifold-based image processing. Propose a new nonlinear dimensionality reduction method with linear projection and diffusion map.

- Super Resolution Imaging for Micro Bubble Localization:

- Study and compare CLEAN, matching pursuit (MP), and Orthogonal Matching pursuit (OMP) algorithms to reconstruct the clear image of the blurred human vascular image by deconvolution.
- Conduct theoretical analysis to justify the equivalence of CLEAN and MP algorithms and demonstrate the advantage of OMP over MP algorithm.

PUBLICATIONS & PREPRINTS

* indicates equal contribution

- [1] **Maohao Shen**, Subhro Das, Kristjan Greenewald, Prasanna Sattigeri, Gregory Wornell, Soumya Ghosh. **Thermometer: Towards Universal Calibration for Large Language Models**. (Under Review)
- [2] Jongha Jon Ryu*, **Maohao Shen***, Soumya Ghosh, Yuheng Bu, Prasanna Sattigeri, Subhro Das, Gregory Wornell. **Improved Evidential Deep Learning via a Mixture of Dirichlet Distributions**. (Under Review)
- [3] Abhin Shah, **Maohao Shen**, Jongha Jon Ryu, Subhro Das, Prasanna Sattigeri, Yuheng Bu, Gregory Wornell. **Group Fairness with Uncertainty in Sensitive Attributes**. (Under Review)
- [4] **Maohao Shen**, Yuheng Bu, Gregory Wornell. **On Balancing Bias and Variance in Unsupervised Multi-Source-Free Domain Adaptation**. (In International Conference on Machine Learning (ICML), 2023.)
- [5] **Maohao Shen**, Soumya Ghosh, Prasanna Sattigeri, Subhro Das, Yuheng Bu, Gregory Wornell. **Reliable Gradient-free and Likelihood-free Prompt Tuning**. (In Conference of the European Chapter of the Association for Computational Linguistics, Findings (EACL), 2023)
- [6] **Maohao Shen**, Yuheng Bu, Prasanna Sattigeri, Soumya Ghosh, Subhro Das, Gregory Wornell. **Post-hoc Uncertainty Learning using a Dirichlet Meta-Model**. (In AAAI Conference on Artificial Intelligence (AAAI), 2023)
- [7] **Maohao Shen***, Bowen Jiang*, Jacky Y. Zhang*, Oluwasanmi Koyejo. **Batch Active Learning from the Perspective of Sparse Approximation**. (In NeurIPS 2022 Workshop on Human in the Loop Learning)
- [8] **Maohao Shen**, Jacky Y.Zhang, Leihao Chen, Weiman Yan, Neel Jani, Brad Sutton, Oluwasanmi Koyejo. **Labeling Sensitive Batch Active Learning for Brain Tumor Segmentation**. In IEEE International Symposium on Biomedical Imaging (ISBI'21), 2021.

[9] Bowen Jiang*, Maohao Shen*. **Dimensionality Reduction Via Diffusion Map Improved with Supervised Linear Projection**. In IEEE International Conference on Image Processing (ICIP'20), 2020.

HONORS AND AWARDS

- Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship, MIT (2021)
- University Honors (top 3%): Bronze Tablet, UIUC (May 2021)
- Edward C. Jordan Award, Dept. of Electrical & Computer Engineering, UIUC (May 2021)
- Henry O. Koehler Merit Scholarship, Dept. of Electrical & Computer Engineering, UIUC (Sept 2020)
- Sargent and Lundy Engineering Scholarship, College of Engineering, UIUC (May 2020)
- Ellery B. Paine Outstanding Junior Award, Dept. of Electrical & Computer Engineering, UIUC (May 2020)
- George A. and Ina M. Zehr Scholarship, Dept. of Electrical & Computer Engineering, UIUC (Sept 2019)
- Dean's List, UIUC (2017-2021)
- James Scholar, UIUC (2017-2021)

MEDIA APPEARANCES

- **Interview by MIT News:** Discuss my research work on uncertainty quantification problems. (Feb 2023) [\[Article Link\]](#)
- **Interview by IBM Research:** Highlight my collaboration with IBM Research to tackle trustworthy learning problems for LLMs. (Jan 2024) [\[Article Link\]](#)

MISCELLANEOUS

- **Skills:** Proficient in Python, PyTorch, C/C++, MATLAB
- **Languages:** Fluent English, Native Mandarin