Lee, Jung Hyun

Address: 102-203, 503 Daedunsan-ro, Jung-gu, Daejeon, South Korea (34987)
E-Mail: onliwad101@gmail.com / LinkedIn: https://www.linkedin.com/in/jung-hyun-lee-9a880b280

EDUCATION

Korea Advanced Institute of Science and Technology (KAIST), Sep. 2019 – Aug. 2021 Daejeon, South Korea
Master of Science in the Graduate School of AI (advisor: Professor Eunho Yang)
- GPA: 3.92 / 4.3
- Thesis: Cluster-Promoting Quantization with Bit-Drop for Minimizing Network Quantization Loss

Pohang University of Science and Technology (POSTECH), Mar. 2011 – Feb. 2019 Pohang, South Korea
Bachelor of Science in Mathematics, minor in Industrial and Management Engineering
- GPA: 3.82 / 4.3 (*Magna Cum Laude*); major GPA: 4.01 / 4.3
- Awarded scholarships including full tuition and additional stipend from Korea Student Aid Foundation (KOSAF)

PUBLICATIONS & ACADEMIC PAPERS (*: Equal Contribution)

Preprints

[8] **Jung Hyun Lee***, June Yong Yang*, Byeongho Heo, Dongyoon Han, Kang Min Yoo. Token-Supervised Value Models for Enhancing Mathematical Reasoning Capabilities of Large Language Models. (Under Review)

[7] **Jung Hyun Lee***, Jeonghoon Kim*, June Yong Yang, Se Jung Kwon, Eunho Yang, Kang Min Yoo, and Dongsoo Lee. LRQ: Optimizing Post-Training Quantization for Large Language Models by Learning Low-Rank Weight-Scaling Matrices. (Under Review)

[6] HyperCLOVA X Team. HyperCLOVA X Technical Report.

Peer-reviewed Articles

[5] Byeonghu Na, Yeongmin Kim, HeeSun Bae, **Jung Hyun Lee**, Se Jung Kwon, Wanmo Kang, Il-chul Moon. Label-Noise Robust Diffusion Models. International Conference on Learning Representations (ICLR), 2024

[4] Jeonghoon Kim*, **Jung Hyun Lee***, Sungdong Kim, Joonsuk Park, Kang Min Yoo, Se Jung Kwon, and Dongsoo Lee. Memory-Efficient Fine-Tuning of Compressed Large Language Models via sub-4-bit Integer Quantization. Neural Information Processing Systems (NeurIPS), 2023.

[3] **Jung Hyun Lee***, Jeonghoon Kim*, Se Jung Kwon, and Dongsoo Lee. FlexRound: Learnable Rounding based on Element-wise Division for Post-Training Quantization. International Conference on Machine Learning (ICML), 2023

[2] Kyung-su Kim*, **Jung Hyun Lee***, and Eunho Yang. Compressed Sensing via Measurement-Conditional Generative Models. IEEE Access, 2021 (SCI)

[1] **Jung Hyun Lee***, Jihun Yun*, Sung Ju Hwang, and Eunho Yang. Cluster-Promoting Quantization with Bit-Drop for Minimizing Network Quantization Loss. IEEE/CVF International Conference on Computer Vision (ICCV), 2021

RESEARCH & WORK EXPERIENCE

NAVER Cloud, Mar. 2022 – Present Seongnam, South Korea
Research Scientist, HyperCLOVA X Team
- Devised a new post-training weight-rounding mechanism, FlexRound [3] to flexibly quantize pre-trained weights of encoder-only and decoder-only language models based on individual characteristics of each weight
- Invented a parameter-efficient and quantization-aware adaption technique, PEQA [4] that fine-tunes only quantization step sizes of a quantized large language model (LLM) to reduce memory usage during fine-tuning and accelerate the inference latency
- Proposed a new post-training weight quantization method for LLMs, LRQ [7] that learns low-rank weight-scaling matrices in a block-wise manner to boost the generalization capability of quantized LLMs

- Introduced the token-supervised value model (TVM) [8], which is a new token-level verifier trained to estimate the probability of an intermediate reasoning path being on a promising track toward the correct final answer

Samsung Research, Jul. 2021 – Mar. 2022 Seoul, South Korea
Software Engineer, Data Research Team

- Had programming training in algorithms and data structures as a newly-hired employee and successfully completed the training course by earning its own programming certification
- Analyzed customers' buying behavior patterns, such as purchase frequency, time and occasion; sorted out loyal customers and recommended brand-new electronic products to them

Machine Learning and Intelligence Laboratory, KAIST, Apr. 2019 – Aug. 2019 Daejeon, South Korea
Research Intern (advisor: Professor Eunho Yang)

- Conducted preliminary research into the impact of neural network pruning on the interpretability of neural networks via Layer-wise Relevance Propagation
- Implemented recent algorithms proposed in deep learning and machine learning papers, reproduced the experimental results, and brainstormed how to improve those algorithms for performance enhancement

PROFESSIONAL SERVICE

REVIEWER: NeurIPS (2022-2024), ICLR (2024), ICML (2024), ACL (2024), AAAI (2025)

EXTRACURRICULAR ACTIVITIES

- Led a student mentor and the ambassador group at POSTECH and helped freshmen's adaptation; organized orientations, taught calculus and applied linear algebra, while attending conferences and sessions to introduce educational infrastructure of POSTECH
- Acted as a captain and playing a coach for POSTECH baseball club; won the 2nd prize in the university competition

OTHER INFORMATION

- Worked as an instructor at Army Infantry School for mandatory military service (Jan. 2013 – Oct. 2014); awarded the commendation medal in recognition of effective leadership
- Awarded the TOP 2 in the research track at the N INNOVATION AWARD 2023, an internal excellence in technology awards ceremony hosted by NAVER