Dongjun Lee

Personal Website: Link
akm5825@naver.com

Objective

I am focused on enhancing user personalization and experience through machine learning and deep learning techniques. I seek opportunities to collaborate with forward-thinking teams where I can make meaningful contributions to real-world applications.

Education

Sungkyunkwan University (GPA: 4.29/4.50) Feb. 2022 – Feb. 2024
M.S. *in Electrical and Computer Engineering* *Suwon, Korea*

Daegu Catholic University (GPA: 4.29/4.50) Mar. 2016 – Feb. 2022
B.S. *in Computer Engineering* *Daegu, Korea*

Experience

AI Scientist Mar. 2024 – Present
Maum AI Inc. *Seongnam, Korea*

- Fine-tuned large language models (LLMs) for various NLP applications.
- Deployed and optimized small language models for on-device inference.

Publications

C³: Capturing Consensus with Contrastive Learning in Group Recommendation
Soyoung Kim, Dongjun Lee, and Jaekwang Kim
Under review

Hierarchical Contrastive Learning with Multiple Augmentations for Sequential Recommendation
Dongjun Lee, Donggeun Ko, and Jaekwang Kim
In Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing (Oral)

Debiasing Classifiers by Amplifying Bias with Latent Diffusion and Large Language Models
Donggeun Ko, Dongjun Lee, Namjun Park, Wonkyeong Shim, and Jaekwang Kim
In Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing (Poster)

DiffInject: Revisiting Debias via Synthetic Data Generation using Diffusion-based Style Injection
Donggeun Ko, Sangwoo Jo Dongjun Lee, Namjun Park, and Jaekwang Kim
In IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2024 Workshop in SynData4CV, 2024

Retrieval-Based Disease Prediction for Myocardial Injury after Noncardiac Surgery:Leveraging Language Models as Diagnostic Tools
Namjun Park, Donggeun Ko, Dongjun Lee, San Kim, and Jaekwang Kim
In AAAI 2024 Spring Symposium on Clinical Foundation Models, 2024

Elevating CTR Prediction: Field Interaction, Global Context Integration, and High-Order Representations
Sojeong Kim, Dongjun Lee, and Jaekwang Kim
In Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing (Poster)

How Important is Periodic Model Update in Recommender System?
Hyunsung Lee, Sungwook Yoo, Dongjun Lee, and Jaekwang Kim
In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval

AmpliBias: Mitigating Dataset Bias through Bias Amplification in Few-Shot Learning for Generative Models
Donggeun Ko, Dongjun Lee, Namjun Park, Kyoungrae Noh, Hyeonjin Park, and Jaekwang Kim
In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management

Self-Interactive Attention Networks via Factorization Machines for Click-Through Rate Prediction
Dongjun Lee, Hyunsung Lee, and Jaekwang Kim
In Proceedings of the 24th International Symposium on Advanced Intelligent Systems

Research Projects

Serving on-device small language models
AI Researcher — *Maum AI*

- Supporting variational language models for optimized inference on NPU.
- Fine-tuned the small language model for specific purposes.
- Constructed instruction-following datasets.

Building Korean reasoning models
AI Researcher — *Maum AI*

- Trained language models to equip with reasoning capabilities.
- Enabled models to think like "o1" and respond in Korean.
- Modeled the reward as a scalar value under the Bradley-Terry model.

Knowledge Distillation
AI Researcher — *Maum AI*

- Researched a knowledge distillation framework inspired by Reinforcement learning approaches.
- Distilled the teacher's knowledge to student model on sentence space.

- Improved the student model's performance through weak-to-strong generalization using both strong and weak models.

Expanding the vocabulary of language models
AI Researcher *Maum AI*
- Extended the vocabulary of English-centric LLMs to include Korean tokens.
- Improved performance on Korean benchmarks while minimizing loss on English tasks.

Domain Adaptation
AI Researcher *Maum AI*
- Implemented a domain specific model for business purpose.
- Trained a domain-specific language model through the lens of data composition and training schemes.

Awards & Honors

Best Paper Award
The 24th International Symposium on Advanced Intelligent Systems *2023*

Summa Cum Laude
Department of Computer Science, Daegu Catholic University *2022*

Skills

Programming Languages: Python (intermediate)
Frameworks & Library: Pytorch, Transformers
Domains: Recommendation Systems, Computer Vision, Natural Language Processing
Languages: Korean (Native), English (intermediate)