Minseok Kim

Ph.D. in Data Science [Google Scholar]
Applied Scientist II at Amazon AGI
minseokkim0630@gmail.com http://minseokkim.net/

RESEARCH INTERESTS

- Large language model (LLM) based conversational AI
- Large scale information retrieval and recommender system
- Trustworthy and real-world ML/AI challenges

WORK EXPERIENCE

Applied Scientist II at Amazon AGI - Seattle, WA. Oct 2022 –
- Optimize and evaluate LLM for tier-1 conversational AI service
- Design and deploy multi-language/-locale model for ranking user request handlers
- Design and develop an end-to-end language modeling pipeline for user intent classification

Applied Scientist Ph.D Internship at Amazon Alexa AI - Seattle, WA. Sep 2021 – Dec 2021
- Project: Debiasing Neighbor Aggregation for Graph Neural Network in Recommender Systems
- Manager: Sungjin Lee (Amazon Principal Scientist)

EDUCATION

Korea Advanced Institute of Science and Technology - Daejeon, Korea
- Ph.D., Graduate School of Data Science Sep 2018 – Aug 2022
- Adviser: Jae-Gil Lee
- Thesis: Meta-Learning for Recommender Systems

Korea Advanced Institute of Science and Technology - Daejeon, Korea
- M.S., Graduate School of Knowledge Service Engineering Sep 2016 – Aug 2018
- Adviser: Jae-Gil Lee
- Thesis: Temporal Interval Refinement for Point-of-Interest Recommendation

Hanyang University - Seoul, Korea
- B.S. in School of Computer Science and Engineering Mar 2013 – Aug 2016
- B.S. in College of Policy Science
- GPA: 4.21/4.50 (Graduated *Cum Laude*)
- Early graduation of excellent students

AWARDS & SCHOLARSHIPS

- Inseoh scholarship May 2022
- Outstanding researcher award, KAIST Institute Dec 2020
- Outstanding researcher award, KAIST KSE Dec 2020
- Qualcomm innovation award Dec 2017
- National scholarship for graduate studies, Korea Student Aid Foundation 2016 – 2022
- Scholarship for volunteers, Pine Tree Foundation 2014 – 2016
- Full academic scholarship for gifted, Hanyang University 2013 – 2016

PUBLICATIONS
[C]: CONFERENCE
[J]: JOURNAL

[C15] Xu, D., Qiu, L., **Kim, M.**, Ladhak, F., and Do, J., "Aligning Large Language Models via Fine-grained Supervision," Preprint 2024.
[C14] Shin, Y., **Kim, M.**, Oh, J., Do, J., Taghavi, T., and Xu, P., "Less-loss Prompt Compression via Entropy-Based Word Addition," Preprint 2024.
[C13] Kim, J., Toikkanen, M., Bae, S., **Kim, M.**, and Jung, H., "RepAugment: Input-Agnostic Representation-Level Augmentation for Respiratory Sound Classification," Preprint 2024.
[C12] Song. H., **Kim, M.**, and Lee, J., "Toward Robustness in Multi-label Classification: A Data Augmentation Strategy against Imbalance and Noise," AAAI 2024 (top conference).
[C11] **Kim, M.**, Oh, J., Do, J., and Lee, S., "Debiasing Neighbor Aggregation for Graph Neural Network in Recommender Systems," CIKM 2022.
[J2] Song, H., **Kim, M.**, Park, D., and Lee, J., "Learning from Noisy Labels with Deep Neural Networks: A Survey," TNNLS 2022 (SCI Expanded, impact factor: 10.451)
[C10] **Kim, M.**, Song, H., Shin, Y., Park, D., Shin, K., and Lee, J., "Meta-Learning for Online Update of Recommender Systems," AAAI 2022 (top conference)
[C9] Kim, D., Min, H., Nam, Y., Song, H., Yoon, S., **Kim, M.**, and Lee, J., "COVID-EENet: Predicting Fine-Grained Impact of COVID-19 on Local Economies," AAAI 2022 (top conference).

[C8] Park, D., Song, H., **Kim, M.**, and Lee, J., "Task-Agnostic Undesirable Feature Deactivation Using Out-of-Distribution Data," NeurIPS 2021 (top conference).
[C7] Song, H., **Kim, M.**, Park, D., and Lee, J., "Robust Learning by Self-Transition for Handling Noisy Labels," KDD 2021 (top conference, full/oral paper, research track).
[C6] **Kim, M.**, Song, H., Kim, D., Shin, K., and Lee, J., "PREMERE: Meta-Reweighting via Self-Ensembling for Point-of-Interest Recommendation," AAAI 2021 (top conference).
[C5] Song, H., **Kim, M.**, Kim, S., and Lee, J., "Carpe Diem, Seize the Samples Uncertain "at the Moment" for Adaptive Batch Selection," CIKM 2020 (full/oral paper).
[J1] Song, H., Kim, S., **Kim, M.**, and Lee, J., "Ada-Boundary: Accelerating DNN Training via Adaptive Boundary Batch Selection," Machine Learning, Vol. 109, No. 9, pp. 1837 – 1853, Sep. 2020 (SCI Expanded, impact factor: 2.672). This paper was presented at the journal track of ECML-PKDD 2020.
[C4] **Kim, M.**, Kang, J., Kim, D., Song, H., Min, H., Nam, Y., Park, D., and Lee, J., "Hi-COVIDNet: Deep Learning Approach to Predict Inbound COVID-19 Patients and Case Study in South Korea," KDD 2020 (top conference, full/oral paper, AI for COVID track).
[C3] Song, H., **Kim, M.**, Park, D., and Lee, J., "How Does Early Stopping Help Generalization against Label Noise?" ICML Workshop 2020.
[C2] Park, D., Song, H., **Kim, M.**, and Lee, J., "TRAP: Two-level Regularized Autoencoder-based Embedding for Power-law Distributed Data," TheWebConf 2020 (top conference, full/oral paper).
[C1] Song, H. **Kim, M.**, and Lee, J., "SELFIE: Refurbishing Unclean Samples for Robust Deep Learning," ICML 2019 (top conference, full/oral paper).

PATENTS

[P6] Lee, J, **Kim, M.**, Kang, J., Kim, D., Song, H., Min, H., Nam, Y., and Park, D., "Method and apparatus for predicting imported infectious disease information based on deep neural networks" US Patent Registration No: US11557401B2, Jan. 17, 2023.

[P5] Lee, J, **Kim, M.**, Kang, J., Kim, D., Song, H., Min, H., Nam, Y., and Park, D., "Method and Apparatus for Predicting Confirmed Patients of Infectious Disease Based on Deep Neural Networks" Korean Patent Registration No: 10-2349270-0000, Jan. 05, 2022.

[P4] Lee, J., Kang, J., **Kim, M.** and Lee, J., "Trajectories Embedding Method for Deep Learning and Route Prediction Method Using the Same," Korean Patent Application No: 10-2020-0179620, Dec. 21, 2020.

[P3] Lee, J., Song, H., and **Kim, M.**, "System and Method of Adaptive Bach Selection for Accelerating Deep Neural Network Learning based on Data Uncertainty," Korean Patent Application No: 10-2020-0133132, Oct. 15, 2020.

[P2] Lee, J., Moon, H., Song, H., **Kim, M.**, and Kim, S., "System and Method for Accelerating DNNs Training via Adaptive Batch Selection," Korean Patent Application No: 10-2020-0044159, Apr. 10, 2020.

[P1] Lee, J. and **Kim, M.**, "Apparatus and Method for Recommending Location," Korean Patent Registration No: 10-2114467-0000, May 18, 2020.

TEACHING EXPERIENCE

- AI Computer Vision (KAIST Dept. of Mathematical Sciences, Korea): Summer 2022
- AI College (Seocho-gu office, Korea): Summer-Winter 2019, Spring-Fall 2021, Spring-Fall 2022
- Deep Learning (KAIST IT academy): Winter 2019, Summer 2021, Winter 2021, Summer 2022
- AI Lecture Materials Development Team Leader (KAIST IT Academy): Winter 2020
- Deep Learning Course Mentor (DSME): Winter 2020
- Machine Learning Course Mentor (DSME): Winter 2020
- AI Program (Hankook Tire): Fall 2019
- Data Processing & Visualization (KAIST IT academy): Summer 2019
- KAIST Advanced AI Academy (LG): Spring 2019
- Data Mining and Knowledge Discovery (KSE525 lecture TA, KAIST): Spring 2019
- Analytical Methodologies for Big Data (KSE526 lecture TA, KAIST): Fall 2018, 2020
- Big Data Professional Course (KB bank group): Summer 2017

RELEASED DATASET

- Animal-10N: A real-world noisy dataset of human-labeled online images for 10 animals. [ICML 2019]

ACTIVITIES

- Graduate school representative (Jan 2020 – Sep 2021)
- Reviewer: ICLR, TNNLS, AAAI, NeurIPS, ICML, ECCV, DKE
- Session Chair: CIKM 2022