SUMYEONG AHN

428 S Shaw Ln # 3312, East Lansing, MI, United States, 48824

(+1)517-940-5197, *Firstname* (at) msu (dot) edu

WORK EXPERIENCE

Postdoc Researcher   Sep. 2023 - Current

Michigan State University, United States (MSU)

Dept. of Computer Science and Engineering (CSE)

Host: Prof. Jiayu Zhou

EDUCATION

Ph. D.   Sep. 2017 - Aug. 2023

Korea Advanced Institute of Science and Technology, S.Korea

Kim Jaechul Graduate School of Artificial Intelligence (KAIST AI)

Thesis: Mitigating Dataset Bias for Robust Deep Learning:
                                        From Clean dataset to the Practical Noisy Dataset

Advisor: Prof. Seyouong Yun (KAIST AI), Prof. Yung Yi (KAIST EE)

Master of Science   Mar. 2015 - Aug. 2017

Korea Advanced Institute of Science and Technology, S.Korea

Department of Electrical Engineering (KAIST EE)

Thesis: Multi-armed Bandit Problem with Intra- and Inter- Correlations

Advisor: Prof. Yung Yi (KAIST EE)

Bachelor of Science   Mar. 2011 - Feb. 2015

Korea University, S.Korea

School of Electrical Engineering

Advisor: Taewoong Yun (KU EE)

RESEARCH INTERESTS

Current Interests

- Vision Language Models [Preparing]

- Machine Learning in Medical Domain [Under review (P1)]

- Efficient Large Language Models [EMNLP'23]

- Noisy label [AAAI'23, Under review (P2), ICLRW'23]

- Dataset Bias (a.k.a Spurious correlation) [ICLR'23, AAAI'23, Under review (P3)]

Previous Interests

- Class imbalance [ICLR'23 (Spotlight)]

- Unsupervised Domain Adaptation with GAN [ICPR'20]

- Stochastic Multi-armed Bandits with Additional Conditions [SIGMETRICS'18]

- AI for Computer Networking Systems [SIGMETRICS'18, MobHoc'21]

PUBLICATIONS

Preprint or Preparing   ∗ : Equal Contribution, _ : Equal Advising

[P4] **Sumyeong Ahn**, Siqi Liang, and Jiayu Zhou. Prompt learning with noisy labels (modified for anonimity). a. (Under Review).

[P3] **Sumyeong Ahn** and Se-Young Yun. Orbis: Open dataset can rescue you from dataset bias problem. a. (Under review).

[P2] **Sumyeong Ahn**, Sihyeon Kim, Jongwoo Ko, and Se-Young Yun. Fine-tuning pre-trained models for robustness under noisy labels. b. (Under review).

[P1] Xiaodan Zhang, Sandeep Vemulapalli, Nabasmita Talukdar, **Sumyeong Ahn**, Jiankun Wang, Han Meng, Sardar Mehtab Bin Murtaza, Aakash Ajay Dave, Dmitry Leshchiner, Dimitri F. Joseph, Martin Witteveen-Lane, Dave Chesla, Jiayu Zhou, and Bin Chen. Large language models in medical term classification and unexpected misalignment between response and reasoning. a. (Under Review).

Conference

[C9] Jihwan Bang, **Sumyeong Ahn**, and Jae-Gil Lee. Active prompt learning in vision language models. (CVPR'24 To appear).

[C8] Xiodan Zhang, Nabasmita Talukdar, Sandeep Vemulapalli, **Sumyeong Ahn**, Jiankun Wang, Han Meng, Sardar Mehtab Bin Murtaza, Dmitry Leshchiner, Aakash Ajay Dave, Dimitri F. Joseph, Martin Witteveen-Lane, Dave Chesla, Jiayu Zhou, and Bin Chen. Comparison of prompt engineering and fine-tuning strategies in large language models in classification on clinical notes. b. (AMIA'23 Informatics Summit).

[C7] Jongwoo Ko, Seungjoon Park, Yujin Kim, <u>**Sumyeong Ahn**</u>, Du-Seong Chang, Euijai Ahn, and <u>Se-Young Yun</u>. Nash: A simple unified framework of structured pruning for accelerating encoder-decoder language models. (EMNLP'23, Findings).

[C6] **Sumyeong Ahn***, Seongyoon Kim*, and Se-Young Yun. Mitigating dataset bias by using per-sample gradient. 2023a. (ICLR'23).

[C5] **Sumyeong Ahn***, Jongwoo Ko*, and Se-Young Yun. CUDA: Curriculum of data augmentation for long-tailed recognition. 2023b. (ICLR'23 **Spotlight**).

[C4] **Sumyeong Ahn** and Se-Young Yun. Denoising after entropy-based debiasing a robust training method for dataset bias with noisy labels. b. (AAAI'23).

[C3] Sangwoo Moon, **Sumyeong Ahn**, Kyunghwan Son, Jinwoo Park, and Yung Yi. Neuro-dcf: Design of wireless mac via multi-agent reinforcement learning approach. (ACM MobiHoc'21).

[C2] Tran Hai H, **Sumyeong Ahn**, Taeyoung Lee, and Yung Yi. Enlarging discriminative power by adding an extra class in unsupervised domain adaptation. (ICPR'21).

[C1] Donggyu Yun, **Sumyeong Ahn**, Alexandre Proutiere, Jinwoo Shin, and Yung Yi. Multi-armed bandit with additional observations. (ACM SIGMETRICS'18).

Workshop

[W3] Jongwoo Ko*, **Sumyeong Ahn***, and Se-Young Yun. Efficient utilization of pre-trained model for learning with nosiy labels. 2023. (ICLR'23 Trustworthy Workshop).

[W2] **Sumyeong Ahn***, Jongwoo Ko*, and Se-Young Yun. CUDA: Curriculum of data augmentation for long-tailed recognition. a. (NeurIPS ML Safety Workshop, NeurIPS Distribution Shifts Workshop).

[W1] **Sumyeong Ahn***, Seongyoon Kim*, and Se-Young Yun. Mitigating dataset bias by using per-sample gradient. b. (NeurIPS ML Safety Workshop, NeurIPS Distribution Shifts Workshop).

Journal

[J1] Donggyu Yun, **Sumyeong Ahn**, Alexandre Proutiere, Jinwoo Shin, and Yung Yi. Multi-armed bandit with additional observations. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 2(1):1–22, 2018.

Patent

[J1] Nansol Seo, Jaemoon Lee, Jaeha Ahn, **Sumyeong Ahn**, Suho Shin, Yoonpyo Koo, and Yung Yi.

Method for controlling multiple uavs based on multi-hop wireless mesh networks, September 3 2019. KOR patent number: 10-2019874-0000.

MISC (Korean conference and journal)

[J1] Sangmin Bae, Taehyeon Kim, **Ahn, Sumyeong**, Sangmook Kim, Jongwoo Ko, and Se-Young Yun. Client sampling algorithm in federated learning via combinatorial averaging and multi-armed bandits. *KIISE*.

[J2] Yoonpyo Koo, **Sumyeong Ahn**, Kyounghwan Son, Suho Shin, Jeonghun Yu, Jaesin Kim, and Yung Yi. An implementation of multi-hop voice communication system using drones. *KICS*.

PROJECT

Leading the project as a student leader

- Developing NAS (Neural Architecture Search) algorithm for detecting layer-wise communication network failures in multi-network-layer datasets.

KT (08/2021-08/2022)

- **Keyword:** Neural Architecture Search (NAS), Communication Network, failure detection

- **Abstract:** The communication networking hardware is often widely dispersed geographically. However, when it generates failure information, a quick resolution is required to mitigate the failure, even if the failure is caused by temporary malfunctioning. Additionally, the information in the networking system continuously changes, such as the occurrence of new log types, and accumulates over time. In this project, we aim to create a Neural Architecture Search (NAS) model using networking log data, including tabular and topological data, to alleviate the burden on AI engineers tasked with creating a model to determine whether a breakdown requires immediate employee intervention.

- Implementing an Application Service for Open/Close Prediction of Commercial Stores using Power Consumption Data

KEPRI (04/2018-12/2019)

- **Keyword:** Predicting Open/Close Status using 1D-CNN and RNN on Time-series Power Consumption Data

- **Abstract:** To enhance customer convenience, several platform providers such as Google and Naver offer Open/Close information in addition to other details about commercial stores. However, inaccuracies in the information, caused by unusual events such as vacations, can make users uncomfortable. To overcome this issue, we have developed a method to predict Open/Close status using time-series data obtained from smart meters. Our aim is to provide a more reliable and convenient service for users. Please follow the link below to access our service, **POWERON**.

- Versatile Network System Architecture for Multi-dimensional Diversity

IITP (04/2016-08/2019)

- **Keyword:** Edge Networking, Fog Computing, FogOS, Matching between resources and requests

- **Abstract:** To leverage the increasing availability of high-speed terminal devices and enhance the performance of cloud computing, edge networking (also referred to as Fog Computing) is being developed. One of the ways to improve performance is by utilizing a module that can efficiently match requests with a set of resources, since resources in edge devices are not as powerful as those in cloud servers. To facilitate seamless cooperation among edge devices, we have developed FogOS, a platform that manages resources and requests and offers services to enhance the performance of edge networking.

- Designing a Matching Algorithm for Efficient Resource Allocation in Distributed Edge Devices

KISA (06/2015-06/2016)

- **Keyword:** Matching algorithm, FogOS, Matcher

- **Abstract:** We conducted a sub-project that established a connection between resources and requests on edge devices.

- Modeling Multi-Hop Wireless Networks for Military Applications

ADD (09/2015-12/2016)

- **Keyword:** Flying Ad-Hoc Network (FANET), Ground Control Station (GCE), Voice Communication, UAVs, Drone

- **Abstract:** Although the Line of Sight (LoS) signals from flying objects such as UAVs and drones are powerful, they have the limitations of short stay times and high shaking. With the aim of enhancing the success rates of military or disaster missions, we embarked on a project that utilized these drones to create an ad-hoc voice call system. We accomplished this by developing a ground control station that could effectively manage a large number of drones and a voice call system with a range of 1 km.

Take part in the project as a participating Student

- Video Transmission over Vehicular Networks

Bosch (01/2015-06/2015)

- **Keyword:** Vehicular Ad-Hoc Network (VANET), 802.11, WiFi, Video codec, real-time video communication, MPEG, JMPEG

- **Abstract:** In vehicular communication, there is strong linearity among physical communication channels, but significant mobility is also present. To improve the transfer of images from the front car to the rear car, we conducted a study and evaluated several codecs to build a networking protocol. Our findings suggest that it is most effective to communicate directly between cars to minimize re-transmission, as dropped frames are meaningless in real-time video communication. Additionally, we recommend uniformizing the importance of each frame by avoiding the use of i-frames in MPEG, as one informative dropped frame can affect other frames.

AWARDS

- KAIST EE Best Teaching Assistant Award (2019 Fall - Embedded System)

SKILLS

- Programming Languages & Frameworks (Selected)

- Machine learning tools: PyTorch, Tensorflow, Scikit-learn

- Database: MySQL, Oracle Database

- System Programming: Bash, C/C++

- Others: OpenCV, Numpy, Matlab

TEACHING

- Teaching Assistance (@ KAIST)

- EE209 Programming Structures for Electrical Engineering (2015 Spring, 2015 Fall, 2020 Spring)

- EE323 Computer Networks (2016 Spring)

- EE205 Data Structures and Algorithms for Electrical Engineering (2017 Fall)

- HSS190(J) Freshman Seminar 1 (2018 Spring)

- EE414 Embedded Systems (2016 Fall, 2019 Fall)

- AI505 Optimization for AI (2020 Fall)

- AI603 Machine Learning Theory (2021 Spring)

- Mentoring

- Jongwoo Ko, KAIST AI (Efficient LM)

- Seungjoon Kim, KAIST AI (Efficient LLM)

- Sangmin Bae, KAIST AI (Efficient LLM)

- Joonki Kim, KAIST AI (Efficient LLM)

- Sihyeon Kim, KAIST AI (Multi-Modal Learning)

- Siqi Liang, MSU (Federated Learning)

- Jiankun Wang, MSU (LLM Application (Medical))

COURSES

- Machine Learning

- (EE837) Deep Learning for Computer Vision (Prof. Junmo Kim @ KAIST EE)

- (AI604) Deep Learning for Computer Vision (Prof. Jaegul Choo @ KAIST AI)

- (AI612) Machine Learning for Healthcare

- Networking

- (EE655) Economics in Communication Networks

- (EE827) Epidemics and Information Diffusion in Complex Networks

- (EE827) 3G/4G/5G Cellular Network Protocols

- (EE650) Optimization in Communication Networks

- (EE627) Performance Analysis of Communication Networks

- (EE827) Network Science

- (EE528) Engineering Random Processes

- Math

- (MAS241) Analysis 1

- (IE539) Convex Optimization

- System

- (EE516) Embedded Software

- (EE324) Network Programming