Matthew DongKyu Cho

Master's Candidate @ Seoul National University · Causal Data Science

kulupapa1127@snu.ac.kr | umamicode.github.io | github.com/umamicode |

linkedin.com/in/dong-kyu-cho-023259176

Personal Profile

A Master's candidate in the Graduate School of Data Science at Seoul National University, under the supervision of Professor Sanghack Lee. My field of interests include Causal Inference, Representation Learning, and Domain Generalization.

Education

Seoul National University — *Seoul, South Korea*

MSc in Data Science — *March 2021 - Current*

- Research Assistant at Causality Lab
- Advised by Sanghack Lee
- **Courses:** Machine Learning and Deep Learning for Data Science, Big Data and Knowledge Management Systems, Software Platforms for Data Science, Machine Learning for Visual Understanding, Text Analytics and Big Data, Causal Inference for Data Science, Special Topics in Data Science: Meta Learning, Data Science Capstone Project, Dissertation Research, Special Lecture on Data Science

Seoul National University — *Seoul, South Korea*

Bachelor of Arts in Information Science and Culture / Western History (Double Major) — *March 2014 - February 2021*

- Graduated with Distinction (Cum Laude)
- Data-Driven Marketing Project with Local Stitch

Research

Distillation-based Single Source Domain Generalization via Mutual Information Regularization — *Seoul, South Korea*

Seoul National University Causality Lab — *Working Paper (Under Review)*

- **Abstract:** Machine learning frequently suffers from the discrepancy in data distribution, commonly known as domain shift. Single-source Domain Generalization (sDG) is a task designed to simulate domain shift artificially, in order to train a model that can generalize well to multiple unseen target domains from a single source domain. A popular approach is to learn robustness via the alignment of samples generated by data augmentation. However, prior works frequently overlooked what can be learned through such alignment. In this paper, we study the effectiveness of augmentation-based sDG methods by connecting recent identifiability results. We highlight the overlooked issues in using augmentation for OOD generalization and search ways to alleviate them. We introduce a novel sDG method that leverages pretrained models to guide the learning process via a feature-level regularization of mutual information, which we name PROF (Progressive mutual information Regularization for Online distillation of Frozen oracles). PROF can be added to conventional augmentation-based methods to dampen the fluctuation of the OOD performance. We further introduce a data-effective alignment objective as well as a novel augmentation method for fine-grained simulation of domain shift.
- Keywords: Causal Representation Learning, Causal Inference, Domain Generalization, Adversarial Data Augmentation, Knowledge distillation

Projects

Stock Interrelation Research using Keyword and Supply Chain data — *Seoul, South Korea*

Seoul National University and NH Investment & Securities — *July 2021 - August 2022*

- A collaborative research with NH Investment & Securities.
- Built a Domain-adapted Language Model and its applicable pipeline using large scale Financial Corpus.
- Built an Interactive Graph Database using the Bloomberg Supply Chain Data.
- **Technical Skills:** Financial-Domain specific Language Models, Financial Data, Graph Database.
- **Soft Skills:** Teamwork, Presentation skills, Report writing.

Towards Language Models Capable of Causal Reasoning — *Seoul, South Korea*

Seoul National University and LG AI Research — *August 2022 - Present*

- Collaborative research with LG AI Research & UNIST
- A Causal NLP Project Under Progress
- **Technical Skills:** Causal Inference, Large-scale Language Models.

Semi-Supervised Federated Learning with Representations *Seoul, South Korea*

Seoul National University - Class Project *March 2022 - July 2022*

- A Class Project for *Special Topics in Data Science: Meta Learning*.
- A Semi-Supervised Approach towards federated learning using Contrastive Learning methods.
- **Technical Skills:** Meta Learning, Self-Supervised Learning, Federated Learning.
- **Soft Skills:** Report writing, Presentation Skills.

Work Experience

LG AI Research *Seoul, South Korea*

Research Intern *July 2023 - Present*

- Causality-Driven Prediction & Optimization for Decision-Making

VAIV Company (former *DaumSoft*) *Seoul, South Korea*

Data Marketing Intern *Sept 2020 - Aug 2021*

- Data-Driven Marketing projects for corporate clients
- Marketing based on Sentiment Analysis of the Korean Social Media.

Republic of Korea Armed Forces *Icheon, South Korea*

Military Interpreter *March 2016 - December 2017*

- Military Interpreter for the Republic of Korea Armed Forces.
- **Technical Skills:** Korean-English Interpretation

Skills

Programming	Python, R, SQL.
Miscellaneous	Linux, Shell (Bash/Zsh), Neo4j, LaTeX(Overleaf/R Markdown), Tableau, Git.
Soft Skills	French, Korean, English-Korean Interpretation

Languages

English	Professional proficiency/ TOEFL 118 (2020.08), 116 (2022.08)
Korean	Native proficiency
French	Intermediate proficiency/ DELF B1

References available upon request.