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 a Korean PropTech startup (Local Stitch)

Work Experience

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

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.

Skills

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

Achievements

2022 **X /340,** Graduate Record Examinations (GRE) *South Korea*

Research

Revisiting Style-Content Disentanglement for Domain Shift from Causal Invariance Principle *Seoul, South Korea*

Seoul National University Causality Lab *Working Paper*

- Masters Dissertation Paper
- **Abstract:** The empirical success of Self-Supervised Learning (SSL) entices researchers to explicate the underlying mechanism of SSL. Among such attempts, recent studies based on the idea of *Causal Invariance* provides a firm theoretical explanation for Self-Supervised Learning. *If* Self-Supervised Learning works on top of a causal mechanism, *then* we may as well devise a new method for the use of *SSL* under domain shifting conditions. In this paper, we provide a Domain Generalization & Adaptation Method using the Disentanglement of *Style* (Variant Features) and *Content* (Invariant Features). We also present a Domain-Agnostic Self-Supervised learning method based on the aforementioned idea of Causal Invariance.
- **Keywords:** Causal Inference, Domain Generalization, Self-Supervised Learning.

Languages

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

References available upon request.