Sungjun Han

Email: sabiosj@gmail.com **Linkedin**: hansungj **Github**: hansungj

Education

M.S., Computational Linguistics, Institute for NLP (IMS) - Universität Stuttgart Nov 2020 - May 2023

Stuttgart, Germany

- GPA of 1.2/5.0 (1.0 highest)
- Thesis supervised by Prof. Dr. Sebastian Pado on the topic "Compositional Generalization through Learning to In-context Learn"

BASc, Engineering Science, University of Toronto Sep 2015 - Apr 2020

Toronto, Canada

- Specialization in Mathematics, Statistics and Finance option
- Graduated with Honours and GPA of 3.5/4.0
- Thesis supervised by Prof. Yuri Lawryshyn under CMTE

Publications

Sungjun Han (2023). *Compositional Generalization through Learning to In-context Learn*. [Master's thesis, University of Stuttgart]. Stuttgart.

Sungjun Han, Deepak Baby, & Valentin Mendelev (2022). *Residual adapters for targeted updates in RNN-transducer based speech recognition system*. In *SLT 2022*. [link]

Experience

Applied Scientist Intern, Amazon Alexa Nov 2021 - Apr 2022

Aachen, Germany

- 5 month full-time research internship supervised by Dr. Deepak Baby and Dr. Valentin Medelev
- Worked on incremental learning for recurrent neural network transducer (RNN-T) for automatic speech recognition system using production english-US data
- Published the main research results at the conference 2022 IEEE Spoken Language Technology Workshop (SLT22)

Research Intern, Mercedes-Benz Speech Technology May 2021 - Oct 2021

Sindelfingen, Germany

- 5 month part-time research internship supervised by Dr. Stefan Ultes
- Worked on developing generative language models for efficient controllable generation for both task-oriented and chit-chat dialogue systems

Software Engineer - Thesis, Woleseley Sep 2019 - Apr 2020

Toronto, Canada

- Year-long Bachelor Thesis supervised by Prof. Dr. Yuri Lawryshyn under CMTE
- Implemented a deep neural network to predict the vendor's delivery time on orders for better inventory management
- Developed a data pipeline to process and prepare the Company's vendor data for a machine learning model using SQL and Python
- Implemented a character-level language model to process natural language data embeddings for the vendor delivery prediction model

Research Intern, D-LAB May 2019 – Aug 2019

Bangkok, Thailand

- Research internship fully-funded through Engineering Science Research Opportunities Program (ESROP)
- Collected and annotated facial emotion recognition data for the virtual assistant for the elderly project
- Assessed the capability of R-CNN family algorithms for facial emotion recognition in a data-scarce setting using Python, PyTorch, and Tensorflow
- Implemented and tested the capability of Bayesian neural network for gene-expression data using Python and Tensorflow

Data Analyst, Scotiabank May 2018 – Apr 2019

Toronto, Canada

- Year-old internship through Progression Experience Year (PEY) program
- Developed and maintained custom queries from the structured database using SQL
- Built a framework for automating the transfer of thousands of business records in Excel to a GUI data platform using Python and Selenium
- Developed a pipeline to store, process and convert data from a structured database into a required format for database migration in Python, SQL and VBScript

Skills

- **Programming Languages:** Python, C, MATLAB, SQL, R
- **Tools:** Git, PyTorch, Tensorflow, Pandas
- **Deep learning and Machine Learning**
- **Languages:** English (native), Korean (native)

Courses & Projects

Universität Stuttgart

- Reinforcement Learning
- Bayesian Statistics and Probabilistic Machine Learning
- Advanced Machine Learning in Speech Processing and Natural Language
  - Project on studying structural priors in graphical vision&language models [github]
- Computational Linguistic Laboratory
  - Project on further pre-training language models for improvement in NLI [github]

- *Deep Learning in Speech Processing and Natural Language*
  - *1st out of 138 classmates in in-class NLP competition in slot-filling NLU*
- *Grounded Lexical Semantics*
- Language and Vision
  - Study of compositionality in visual-question-answering (VQA) models [github]
- Machine Translation
- Methods in Computational Linguistics - Morphology, Generative Syntax, Formal Semantics
- Speech Recognition and Synthesis

University of Toronto

- Algorithms and Data Structures
- Capstone Design
  - Machine learning pipeline monitoring system in collaboration with a start-up based in Toronto [slides]
- Financial optimization
  - Portfolio optimization using MVO, Robust-MVO, Resampling-MVO, MD-MVO, and CVaR [github]
- Intro to Databases
- Methods of Data Analysis
- Probability and Statistics

Kaggle

- 2020 Jigsaw Multilingual Toxic Comment Classification competition
  - 34th out of 1221 teams through the XLM-RoBERTa zero-shot transfer solution [solution]