Hwijeen Ahn

hwijeenahn@cmu.edu / hwijeen.github.io

Research Interest

Direction-wise: Multilingual NLP, Social NLP
Method-wise: Efficient transfer learning, Linguistic structure in neural models

Education

Carnegie Mellon University — Sep. 2022 - Current
Master in Language Technologies — Pittsburgh, USA
Advisor: Eduard Hovy

Sogang University — Mar. 2018 - Feb. 2021
M.S. in Computer Science and Engineering — Seoul, Korea
CGPA: 4.38 / 4.5
Advisor: Jungyun Seo

Sogang University — Mar. 2011 - Feb. 2018
B.A. in American Culture, Psychology — Seoul, Korea
Summa Cum Laude
2-year leave of absence as military interpreter

Research Experience

Research Engineer — Jan. 2021 - Apr. 2022
NAVER, CLOVA — Seongnam, Korea
- Studied effect of pretraining corpora on in-context learning of GPT-3
- Implemented prompt tuning methods using model and data parallelism
- Developed framework for pretraining and distributing language models

Visiting student — Aug. 2019 - Feb. 2020
Language Technologies Institute, Carnegie Mellon University — Pittsburgh, USA
- Proposed cultural similarity features for cross-lingual transfer
- Devised crosslingual data selection method for offensive language detection
- Attained GPA of 4.00 / 4.33 in graduate level coursework

Research Assistant — Mar. 2018 - Feb. 2021
Sogang Natural Language Processing Lab — Seoul, Korea
- Proposed denoising-based grammar error correction method
- Developed SOTA Korean dependency parser using morpheme information
- Researched method to detect bias in multilingual neural models

Publications

Seongjin Shin* , Sang-Woo Lee*, Hwijeen Ahn, Sungdong Kim, HyoungSeok Kim, Boseop Kim, Kyunghyun Cho, Gichang Lee, Woomyoung Park, Jung-Woo Ha, Nako Sung. 2022. On the Effect of Pretraining Corpora on In-context Few-shot Learning by a Large-scale Language Model. *Proceedings of the 2022 Conference on North American Chapter of the Association for Computational Linguistics (NAACL).* [pdf]

Hwijeen Ahn*, Jimin Sun*, Chan Young Park*, Yulia Tsvetkov, David R. Mortensen. 2020. Cross-Cultural Similarity Features for Cross-Lingual Transfer Learning of Pragmatically Motivated Tasks. *Proceedings of the 2021 Conference on European Chapter of the Association for Computational Linguistics (EACL).* [pdf]

Hwijeen Ahn*, Jimin Sun*, Chan Young Park*, Jungyun Seo. 2020. NLPDove at SemEval-2020 Task 12: Improving Offensive Language Detection with Cross-lingual Transfer. In *Proceedings of the Fourteenth Workshop of Semantic Evaluation (SemEval).* [pdf]

Hwijeen Ahn, Minyoung Seo, Chanmin Park, Juae Kim, Jungyun Seo. 2019. Extensive Use of Morpheme Features in Korean Dependency Parsing. In *Proceedings of the IEEE International Conference on Big Data and Smart Computing (BigComp).*

Honors and Awards

- N Innovation Award (Research and development track), NAVER CLOVA, 2021. Won $30,000 as team for pretraining Korean GPT-3 and developing tune-as-a-service.
- N Innovation Award (Research and development track), NAVER CLOVA, 2021. Participated in efficient prompt tuning of Korean GPT-3 for CareCall project.
- N Innovation Award (Sharing lessons learned track), NAVER CLOVA, 2021. Won $3,000 for sharing experience in developing a unified framework for language models.
- Winner, NAVER AI RUSH, 2020. Achieved competitive performance in grammar error correction competition and spam mail detection, winning $18,000.
- Gold Prize, NLP competition by National Institute of Korean Language, 2018. Achieved state-of-the-art performance in Korean dependency parsing.
- Popularity Prize, Sogang University Hackathon, 2017
- CORE Scholarship, National Research Foundation of Korea, 2017
- Semester High Honors, Sogang University, 2016,2017
- Jeannine Manuel Award, Ecole Jeannine Manuel, 2008

Academic activities

- Oral presentation on pretraining large scale language models, NAVER DEVIEW, 2021
- Oral presentation on natural language review generation from categorical attributes, BOAZ Big Data Society, 2020

Patent

Korean dependency parser (Patent No. 10-2018-0148817)

Technical Skills

Python stack: PyTorch, Transformers, Pytorch Lightning, Hydra, MegatronLM
Older experience: C, C++
Others: Git, Docker, Vim, Tmux, Shell, Linux

Language

Fluent in **Enlgish**, Native in **Korean**, Intermediate in **French**