Hwijeen Ahn
hahn2@cs.cmu.edu | +1 412-613-9903 | hwijeen.github.io | www.linkedin.com/in/hwijeenahn/

Education

Carnegie Mellon University	Pittsburgh, PA
Master in Language Technologies — *Advisor: Eduard Hovy*	Aug. 2024
6-month visiting student — *Funded by Korean government*	Feb. 2020
Sogang University	Seoul, Korea
M.S. in Computer Science and Engineering — *Advisor: Jungyun Seo*	Feb. 2021
B.A. in Psychology, American Culture — *Summa Cum Laude*	Feb. 2018
2-year leave of absence as military interpreter	

Professional Experience

Apple — Seattle, WA
Machine Learning Engineer Intern — *Mentor: Aaron Jaech* — May. 2023 - Aug. 2023
- Formulated speech language understanding task as combination of discriminative and generative problems and implemented system where each component is trained using synthetic and distantly supervised data
- Reformulated the objective function to reduce redundant generation in decoder, deployed the model with int8 quantization, and batched generation process on server to make inference 2x faster on MPS device
- Developed RESTful API based backend server that runs the generative model in thread-safe manner in real time, applies pre/post processing on the request, logs usage to database

NAVER — Seongnam, Korea
Machine Learning Research Engineer — *Language Representation Team* — Jan. 2021 - May. 2022
- Designed framework for parameter-efficient finetuning of GPT3 based on Ptuning, LoRA and implemented it using Megatron-LM style model parallelism gaining 100+ users company-wide that adapted large language model for their task
- Ran large-scale ablation study on cluster of 1024 NVIDIA GPUS to identify the effect of each billion-token scale domain-specific corpora and their combination on in-context learning performance of 1.3B-sized Korean GPT3 and coauthored NAACL 2022 paper [paper]
- Developed language model library compatible with huggingface ecosystem and used it to pretrain 20+ different versions of BERT, ELECTRA, GPT up to model size of 770M and distributed them company-wise resulting in more than 1000+ downloads [talk]

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. On the Effect of Pretraining Corpora on In-context Few-shot Learning by a Large-scale Language Model. *NAACL 2022.* [paper]

Hwijeen Ahn*, Jimin Sun*, Chan Young Park*, Yulia Tsvetkov, David R. Mortensen. Cross-Cultural Similarity Features for Cross-Lingual Transfer Learning of Pragmatically Motivated Tasks. *EACL 2021.* [paper]

Hwijeen Ahn*, Jimin Sun*, Chan Young Park*, Jungyun Seo. NLPDove at SemEval-2020 Task 12: Improving Offensive Language Detection with Cross-lingual Transfer. *SemEval 2020.* [paper]

Hwijeen Ahn, Minyoung Seo, Chanmin Park, Juae Kim, Jungyun Seo. Extensive Use of Morpheme Features in Korean Dependency Parsing. In *IEEE BigComp 2019.* [patent]

Recent Projects

Photobook: multimodal reference game — *CMU Multimodal ML class* — Jan. 2023 - May. 2023
- Bulit speaker model for dialog based reference resolution game employing prefintuning of 2.7B BLIP-2 on image captioning dataset to improve grounding and task-oriented accuracy [report]

Masking as alternative to finetuning — *Independent Study with Emma Strubell* — Sep. 2022 - Dec. 2022
- Studied masking as efficient alternative to finetuning and found that adapting pretrained models by masking 5% of weights leads to better task transfer compared to finetuning in intermediate task transfer setting [report]

Grammar error correction — *NLP competition by NAVER* — Jul. 2020 - Sep. 2020
- Proposed synthetic data generation strategy based on self-training and backtranslation and pretrained transformer to achieve +5 BLEU score improvement and won $10,000 cash prize [interview]

Skills / Open Source Contribution

Programming languages: Python, C
ML Ecosystem: PyTorch, Transformers([merged PR1], [2], [3]), Datasets, MegatronLM([merged PR]),
Pytorch-Lightning, Numpy, Pandas, Hydra, Scikit-learn, PySpark
Dev environments: Vim, Tmux, Git, Shell, Docker, Databricks