Hyeongu Yun

· AI Scientist @ LG AI Research ·

(+82) 10-8742-9670 | youaredead@snu.ac.kr

Summary

An AI Scientist in LG AI Research with a year experience of training generative large language model with cutting-edge algorithms, curating massive corpora tailored for pre-training to align with specific objectives, and crafting instruction fine-tuning datasets to optimize the model's interactive capability with expected user's behaviors.

Work Experiences

EXAONE Lab. @ LG AI Research — *Seoul, Republic of Korea*

AI Scientist — *Jul. 2022 - Current*

- Developed "in-house" generative decoder-only large language models in multiple sizes and domains.
- Curated extensively massive corpora (up to 2 trillion tokens) from various sources for foundation LLMs.
- Built instruction fine-tuning datasets to make assistant-like generative LLMs.
- Designed suitable benchmarks for LLMs in various domains with specific objectives including human-preference.

Education

Ph.D. in Department of Electrical and Computer Engineering — *Seoul, Republic of Korea*

Seoul National University — *Mar 2015 - Aug 2022*

Advisor: Kyomin Jung, Professor, Seoul National University
Dissertation Title: Hierarchical Context Encoder for Natural Language Processing via Leveraging Contextual Information and Memory Attention

B.S. in Department of Electrical and Computer Engineering — *Seoul, Republic of Korea*

Seoul National University — *Feb 2015*

Internship Experiences

Research Intern — *Oct. 2021 - Apr. 2022*

at Microsoft Research Lab - Asia

· Research on Data augmentation with a context-controlled text generation model.

Research Intern — *Jun. 2017 - Aug. 2017*

at NAVER Papago

· Research on neural machine translation for documents in HTML format. I developed a matching algorithm that finds an appropriate span for HTML tags on the translated documents corresponding to the original documents.

Publications

In-Context Instruction Learning — *preprint*

Seonghyeon Ye, Hyeonbin Hwang, Sohee Yang, Hyeongu Yun, Yireun Kim, and Minjoon Seo — *Under review process*

PR-MCS: Perturbation Robust Metric for MultiLingual Image Captioning — *preprint*

Yongil Kim, Hyeongu Yun, Seunghyun Yoon, Trung Bui, Yerin Hwang, and Kyomin Jung — *Under review process*

Gated Relational Target-aware Encoder and Local Context-aware Decoder for Target-oriented Opinion Words Extraction — *IEEE Access*

Taegwan Kang, Segwang Kim, Hyeongu Yun, Hwanhee Lee, Kyomin Jung — *page 130507-130517, Volume 10, Dec 2022*

Modality Alignment between Deep Representations for Effective Video-and-Language Learning

LREC 2022

Hyeongu Yun, Yongil Kim, Kyomin Jung

Jun 2022, Marseille (France) and Online

Contrastive Learning for Context-aware Neural Machine Translation Using Coreference Information

WMT 2021

Yongkeun Hwang, Hyeongu Yun, Kyomin Jung

Nov 2021, Punta Cana (Dominican Republic) and Online

Pairwise Context Similarity for Image Retrieval System Using Variational Auto-Encoder

IEEE Access

Hyeongu Yun, Yongil Kim, Taegwan Kang, Kyomin Jung

page 34067 - 34077, Volume 9, Feb 2021

Improving Context-Aware Neural Machine Translation Using Self-Attentive Sentence Embedding

AAAI 2020

Hyeongu Yun, Yongkeun Hwang, Kyomin Jung

Jan 2020, New York City (USA)

Efficient Transfer Learning Schemes for Personalized Language Modeling using Recurrent Neural Network

AAAI Workshop on Crowdsourcing, Deep Learning and Artificial Intelligence Agents

Seunghyun Yoon, Hyeongu Yun, Yuna Kim, Gyu-tae Park, Kyomin Jung

Feb 2017, San Francisco, (USA)

Domestic Publications

Improving Video-QA system through Modality Alignment - Awarded as the Best Paper

KCC 2021

Yongil Kim, Hyeongu Yun, Kyomin Jung

Jun 2021

Image Feature Hashing with Pairwise Context Similarity - Awarded as the Best Presentation Paper

KSC 2019

Hyeongu Yun, Taegwan Kang, Kyomin Jung

Dec 2019

Skills & Interests

Programming Skills

Python ★★★, Tensorflow ★★★★, Pytorch ★★★★★, C++ ★★☆