MINSANG KIM

+82 10 4026 4271 ⋄ kmswin7@gmail.com
https://kmswin1.github.io/

RESEARCH INTERESTS

Representation Learning for structured data (graphs) and unstructured data (text) and their applications

EDUCATION

Korea University, Seoul, South Korea *2021 - 2023*
Master of Science in Engineering
Department of Computer Science and Engineering
System Intelligence Group (SING)
Research area : Representation Learning for structured data (graphs) and unstructured data (text)
Advisor : professor Seungjun Baek

Korea University, Seoul, South Korea *2013 - 2019*
Bachelor of Science in Engineering
Department of Computer Science and Engineering
Data mining and visual analytics Lab (DAVIAN) Internship
Research area : Natural Language Processing, Machine Learning
Advisor : professor Jaegul Choo

PUBLICATIONS

Graph Representation Learning

Cluster and Landmark Attributes Infused Graph Neural Networks for Link prediction - **Minsang Kim**, Seungjun Baek (Under review)

Natural Language Processing

- Persona-Knowledge Dialogue Multi-Context Retrieval and Enhanced Decoding Methods - Minsik Oh*, **Minsang Kim*** (*: equal contributions)
  COLING workshop on Customized Chat Grounding Persona and Knowledge (COLING-PKG), 2022, *Oral, Accepted* **PDF**
  Workshop challenge 1st place award (Both retrieval & generation tasks)
- OASYS: Domain-Agnostic Automated System for Constructing Knowledge Base from Unstructured Text - **Minsang Kim***, Sang-Hyun Je*, Eunjoo Park (*: equal contributions)
  ACM SIGKDD Workshop on Mining and Learning with Graphs (KDD-MLG), 2022, *Poster, Accepted* **PDF**

Knowledge Graph

- ComDensE : Combined Dense Embedding of Relation-aware and Common Features for Knowledge Graph Completion - **Minsang Kim**, Seungjun Baek
  International Conference on Pattern Recognition (ICPR), 2022, *Poster, Accepted* **PDF**

AWARDS

- COLING workshop challenges on Customized Chat Grounding Persona and Knowledge, 2022
  COLING-PKG - All shared tasks 1st place award (1.Retrieval & 2.Generation).

WORK EXPERIENCE

SK Telecom, South Korea December 2019 - current
Machine Learning Research Engineer

· Research area : Representation Learning for structured data (graphs) and unstructured data (text). In particular, Large-scale language models.

Fast Campus, South Korea September 2022 - Current
Lecturer

· I lectured on several NLP categories - Machine translation, Sentence classification, Passage retrieval, Dialogue system, Knowledge graph, and Information extraction. **Lecture**

Kakao Enterprise AI Lab, South Korea December 2019 - December 2022
Machine Learning Research Engineer

· Research area : Representation Learning for structured data (graphs) and unstructured data (text)

Kakao Corp. AI Lab, South Korea July 2019 - December 2019
Machine Learning Research Engineer

· Research area : Representation Learning for structured data (graphs) and unstructured data (text)

PROJECTS

2023.01 - current : SK Telecom
Modeling large-scale pre-trained language models. e.g. pre-train, instruction fine-tuning, etc.
Developing *Korean ChatGPT.*

2022.12 - 2023.02 : Eleuther AI
1.POLYGLOT - Modeling Multilingual pre-trained-language models (from data gathering to training).
2.OSLO - Developing distributed training framework.

2021.10 - 2022.12 : Kakao Enterprise
I had developed Open-Domain QA (Passage Retrieval, MultiModal Retrieval), NLU tasks (Assistant, Bot classifier, Intent Classifier, Machine Learning Entity Tagger) for Kakao-i assistant system, and cross-document language modeling. (estimate semantic network and post-train LMs.)

2021.10 - 2022.12 : Kakao Enterprise
I had developed graph-based models such as link prediction, and community detection for network analysis project.

2021.07 - 2021.12 : Kakao Enterprise
I had developed Sequence to sequence model for triple (knowledge) to text generation task.

2020.08 - 2021.12 : Kakao Enterprise
I had developed a knowledge graph auto-construction system without labeled data where our company requires some automatic training service system for B2B business. We also opened Korean text to triple benchmark dataset for contributing KG construction research environment.

2020.05 - 2021.02 : Kakao Enterprise
I developed personalized member ranking system for KakaoWork that B2B messenger service in our company.

2019.07 - 2020.3 : Kakao Corp.
I had developed Daum search engine that is one of the most famous search engine in South Korea contents search ranking system.

2018.07 - 2019.06 : Data visualization and analytics Lab
I participated in lab paper seminar every weeks and researched Machine translation with lab members. (Topic : Personalized Neural Machine translation)

LEADERSHIP EXPERIENCES

2019.07 - current : Kakao Corp & Kakao Enterprise
Lead several projects and team research seminar.
I had led several service projects in my companies. First, I had developed Daum search engine that is oone of the most famous search engines in South Korea. Second, I developed a personalized member ranking system for KakaoWork that B2B messenger service in our company. In addition, I had developed a knowledge graph auto-construction system without labeled data where our company requires some automatic training service system for B2B business. Third, I have developed graph-based models such as link prediction, and community detection for network analysis project (such as how workers in specific company collaborate with, who would collaborate in the future etc.). Finally, I have developed Open-Domain QA for Kakao-i system, and cross-document language modeling (estimate semantic network and post-train LMs).
I also led and planned a team research seminar for reviewing AI conferences. Every member read the paper and prepared for the presentation. During the seminar, each member presented their papers and discussed these papers.

2017.07 - 2017.08 : Draper University, San Mateo & Industrial Technology research institute, Berlin
KIC college R2BD (Research to Business Development) program : Complete Entrepreneurship program.
In the summer of 2017, I attended to KIC college R2BD program, which was participated by several universities in South Korea, and I was on behalf of our school for cultivating entrepreneurship. I and My colleague introduced a pill scanner that classifies what kind of pills by using a deep learning model and then made a business plan for this project. During this program, I learned about teamwork, design-thinking, presentation skills, etc. from various activities. Furthermore, I had progressed our business plan combined with methods that were taught from various activities. Finally, I presented my project result in front of a lot of international students and Tim Draper who is a famous venture capital investor in Silicon Valley.

TECHNICAL STRENGTHS

Research Area	GNN, NLP, KG, Representation Learning
Software, Tools	python(advanced), pytorch(advanced), LATeX(advanced), slurm(advanced)

LANGUAGE PROFICIENCY

- Korean (Native proficiency)
- English (Advanced proficiency)