Jinheon Baek

Updated October 8, 2023

Email: jinheon.baek@kaist.ac.kr **Homepage**: https://jinheonbaek.github.io **Google Scholar**: /JinheonBaek

Research Interest

My research interest lies in the area of machine learning for graphs, languages, and their intersections at scale. Previous work includes modeling interconnected structures and knowledge, and retrieving them to augment language models for natural language applications, such as question answering, dialogue, and internet search.

Education

KAIST Daejeon, Korea
Ph.D. in Artificial Intelligence Mar 2022 – Present
M.S. in Artificial Intelligence Mar 2020 – Feb 2022
Thesis: Toward Accurate Learning of Graph Representations, GPA: 4.22/4.30

Korea University Seoul, Korea
B.S. in Computer Science and Engineering Mar 2016 – Feb 2020
B.E. in Software Technology and Enterprise Program (Interdisciplinary Program)
GPA: 4.40/4.50, GPA of Computer Science and Engineering: 4.48/4.50

Employment

Research Assistant, KAIST (Advisor: Prof. Sung Ju Hwang) Jul 2019 - Present
- Conducting research on machine learning for graphs (representation, reconstruction, generation, and retrieval), and their applications to natural language (knowledge graphs) and computer vision (object graphs).

Research Intern, Microsoft Research (Redmond) Jun 2023 - Sep 2023
- Conducted research on augmenting large language models with personal knowledge for query suggestions.

Applied Scientist II Intern, Alexa AI, Amazon (Cambridge) Aug 2022 - Nov 2022
- Conducted research on knowledge-augmented large language models with retrieval from knowledge graphs.

Chief Technology Officer (CTO), Jobshopper Jul 2018 - Feb 2020
- Co-founded an education-technology startup for student careers, led a research & development group, exited the big-data-based university information platform (Majormap), and now only advising technical directions.

Undergrad. Research Assistant, Korea University (Advisor: Prof. Jaewoo Kang) Mar 2018 - Jun 2019
- Conducted research on representing interconnected knowledge (food ingredients and academic majors) for their recommendations, and research on investigating the spatial reasoning ability of neural models.

Research Intern, Smart Data Research Group, ETRI Jan 2019 - Feb 2019
- Researched an open data distribution platform with the international standard (DCAT), and utilized DevOps tools (Ansible, AWX, and Traefik) to manage nodes on a cloud, in a National AI Research Institute in Korea.

Publications

International Publications (∗: equal contribution; †: corresponding author)

[26] Personalized Knowledge-Augmented Large Language Models for Contextual Search Query Suggestions
Jinheon Baek, Nirupama Chandrasekaran, Silviu Cucerzan, Allen Herring, and Sujay Kumar Jauhar
Under Review

[25] Knowledge-Augmented Language Model Verification
Jinheon Baek, Soyeong Jeong, Minki Kang, Jong C. Park, and Sung Ju Hwang
Empirical Methods in Natural Language Processing (**EMNLP**), 2023

[24] Test-Time Self-Adaptive Small Language Models for Question Answering
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Findings of Empirical Methods in Natural Language Processing **(Findings of EMNLP)**, 2023

[23] Knowledge-Augmented Reasoning Distillation for Small Language Models in Knowledge-Intensive Tasks
Minki Kang, Seanie Lee, Jinheon Baek, Kenji Kawaguchi, and Sung Ju Hwang
Conference on Neural Information Processing Systems **(NeurIPS)**, 2023

[22] Direct Fact Retrieval from Knowledge Graphs without Entity Linking
Jinheon Baek†, Alham Fikri Aji, Jens Lehmann, and Sung Ju Hwang
Annual Meeting of the Association for Computational Linguistics **(ACL)**, 2023

[21] Phrase Retrieval for Open-Domain Conversational Question Answering
with Conversational Dependency Modeling via Contrastive Learning
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Findings of the Association for Computational Linguistics **(Findings of ACL)**, 2023

[20] Knowledge-Augmented Language Model Prompting for Zero-Shot Knowledge Graph Question Answering
Jinheon Baek†, Alham Fikri Aji, and Amir Saffari
Natural Language Reasoning and Structured Explanations Workshop at ACL **(NLRSE @ ACL)**, 2023 **(Best Paper)**
Matching from Unstructured and Structured Data Workshop at ACL **(MATCHING @ ACL)**, 2023 **(Oral)**

[19] Personalized Subgraph Federated Learning
Jinheon Baek*, Wonyong Jeong*, Jiongdao Jin, Jaehong Yoon, and Sung Ju Hwang
International Conference on Machine Learning **(ICML)**, 2023

[18] Realistic Conversational Question Answering with Answer Selection
based on Calibrated Confidence and Uncertainty Measurement
Soyeong Jeong, Jinheon Baek, Sung Ju Hwang, and Jong C. Park
Conference of the European Chapter of the Association for Computational Linguistics **(EACL)**, 2023

[17] Graph Self-supervised Learning with Accurate Discrepancy Learning
Dongki Kim*, Jinheon Baek*, and Sung Ju Hwang
Conference on Neural Information Processing Systems **(NeurIPS)**, 2022

[16] Object Detection in Aerial Images with Uncertainty-Aware Graph Network
Jongha Kim, Jinheon Baek, and Sung Ju Hwang
Visual Object-oriented Learning meets Interaction Workshop at ECCV **(VOLI @ ECCV)**, 2022

[15] Knowledge-Consistent Dialogue Generation with Knowledge Graphs
Minki Kang*, Jin Myung Kwak*, Jinheon Baek*, and Sung Ju Hwang
Knowledge Retrieval and Language Models Workshop at ICML **(KRLM @ ICML)**, 2022

[14] KALA: Knowledge-Augmented Language Model Adaptation
Minki Kang*, Jinheon Baek*, and Sung Ju Hwang
Conference of the North American Chapter of the Association for Computational Linguistics **(NAACL)**, 2022 **(Oral)**

[13] Augmenting Document Representations for Dense Retrieval with Interpolation and Perturbation
Soyeong Jeong, Jinheon Baek, Sukmin Cho, Sung Ju Hwang, and Jong C. Park
Annual Meeting of the Association for Computational Linguistics **(ACL)**, 2022 **(Oral)**

[12] Edge Representation Learning with Hypergraphs
Jaehyeong Jo*, Jinheon Baek*, Seul Lee*, Dongki Kim, Minki Kang, and Sung Ju Hwang
Conference on Neural Information Processing Systems **(NeurIPS)**, 2021

[11] Task-Adaptive Neural Network Retrieval with Meta-Contrastive Learning
Wonyong Jeong*, Hayeon Lee*, Geon Park*, Eunyoung Hyung, Jinheon Baek, and Sung Ju Hwang
Conference on Neural Information Processing Systems **(NeurIPS)**, 2021 **(Spotlight)**

[10] Unsupervised Document Expansion for Information Retrieval with Stochastic Text Generation
Soyeong Jeong, Jinheon Baek, ChaeHun Park, and Jong C. Park
Scholarly Document Processing Workshop at NAACL **(SDP @ NAACL)**, 2021 **(Oral)**

[9] Accurate Learning of Graph Representations with Graph Multiset Pooling
Jinheon Baek*, Minki Kang*, and Sung Ju Hwang
International Conference on Learning Representations **(ICLR)**, 2021

[8] Exploring The Spatial Reasoning Ability of Neural Models in Human IQ Tests
Hyunjae Kim*, Yookyung Koh*, Jinheon Baek, and Jaewoo Kang
Neural Networks, 2021

[7] Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction
Jinheon Baek, Dong Bok Lee, and Sung Ju Hwang
Conference on Neural Information Processing Systems **(NeurIPS)**, 2020

Domestic Publications, mostly written in Korean

[6] Artificial Intelligence-Based High School Course and University Major
Recommendation System for Course-Related Career Exploration
Jinheon Baek, Hayeon Kim, and Kiwon Kwon
KIPS Transactions on Software and Data Engineering (KTSDE), 2021

[5] Graph Representation Learning with Attention-based Set Pooling
Jinheon Baek, Minki Kang, and Sung Ju Hwang
Conference of Korean Artificial Intelligence Association (CKAIA), 2020 **(Best Paper)**

[4] Embedding Academic Majors and Lectures for Analyzing Departments in University
Jinheon Baek, Gwanghoon Jang, Soyeong Jeong, Donghyeon Park, Kiwon Kwon, and Jaewoo Kang
Korea Computer Congress (KCC), 2019 **(Oral)**

[3] Embedding Food Ingredients Based on Chemical Combination in Dense Vector Space
Donghyeon Park, Yonggyu Park, Buru Chang, Jinheon Baek, and Jaewoo Kang
Korea Computer Congress (KCC), 2018 **(Best Paper)**

[2] Multimodal Approach for Blocking Obscene and Violent Contents
Jinheon Baek, Dakyeong Lee, Chaeyeon Hong, and Byeongtae Ahn
Journal of Convergence for Information Technology (JCIT), 7(6), 2017

Thesis

[1] Toward Accurate Learning of Graph Representations
Jinheon Baek
Master's Thesis, KAIST, 2022

Patents

[3] Electronic Apparatus and Method for Controlling Thereof
Sung Ju Hwang, Jinheon Baek, Wonyong Jeong, Hyeonuk Kim, Jiongdao Jin, and Jaehong Yoon
Korean patent number: 10-2022-0175165, filed on Dec 14, 2022

[2] Method for Building an Artificial Intelligence based Research Topics Database
Kiwon Kwon, and Jinheon Baek
Korean patent number: 10-2021-0017765, filed on Feb 08, 2021, and registered on Apr 19, 2022

[1] Method for Consulting on Exploration Activities
based on Target Departments and Curriculum Achievement Standards
Kiwon Kwon, and Jinheon Baek
Korean patent number: 10-2021-0017764, filed on Feb 08, 2021, and registered on Jun 17, 2022

Honors and Awards

Received the Best Paper Award at NLRSE Workshop in ACL 2023	2023
Awarded the ICML Grant for ICML 2023	2023
Awarded the Google Travel Grant for NeurIPS 2022	2022
Selected as One of the Top Reviewers (Top 10%) of NeurIPS 2022	2022
Selected as One of the Highlighted Reviewers (Top 10%) of ICLR 2022	2022
Selected as One of the Best Reviewers (Top 10%) of ICML 2021	2021
Received the Best Paper Award at CKAIA 2020	2020
Awarded the Samsung Dream Scholarship	2016-2020
Received the First Prize in the Graduation Project Competition Among Around 20 Teams, Computer Science and Engineering Department at Korea University	2019
Received the Academic Excellence Award (highest GPA) at Korea University	2019
Received the Second Prize for Excellence in the Microsoft Student Partners Activities, Among Around 10 Teams as a Team Leader, Microsoft Student Partners	2018
Nominated as a Representative of Korean for Excellence in Microsoft Student Partners Activities, Among Around 50 Members on a Global Summit Program, Microsoft Student Partners	2017

Academic Services

Reviewer of **NeurIPS** (Conference on Neural Information Processing Systems)	2021-2023
Reviewer of **ICML** (International Conference on Machine Learning)	2021-2023
Reviewer of **ICLR** (International Conference on Learning Representations)	2022-2024
Reviewer of **EACL** (Conference of the European Chapter of the Association for Computational Linguistics)	2023
Reviewer of **LoG** (Learning on Graphs Conference)	2022-2023
Reviewer of **TMLR** (Transactions on Machine Learning Research)	2022-2023
Reviewer of **TPAMI** (IEEE Transactions on Pattern Analysis and Machine Intelligence)	2023
Reviewer of **TNNLS** (IEEE Transactions on Neural Networks and Learning Systems)	2023
Reviewer of **TETCI** (IEEE Transactions on Emerging Topics in Computational Intelligence)	2023
Reviewer of **Big Data Research**	2023
Reviewer of **NLRSE @ ACL 2023** (Natural Language Reasoning and Structured Explanations Workshop)	2023

Volunteering Experience

Oxford COVID-19 Government Response Tracker (OxCGRT) Jan 2022 - Mar 2023
- Collected systematic data on policy measures that governments took to tackle COVID-19, aiding decision-makers and citizens in understanding governmental responses over 180 countries. Research papers: [a, b].

WoonWha - Education Volunteer Club, Korea University Jul 2016 - Jul 2017
- Volunteered for 10 middle school students vulnerable to education twice a week, about 300 hours in total.

5th Volunteering Program, Samsung Dream Scholarship Foundation Jan 2017
- Volunteered for children in Cambodia by teaching students and exchanging cultures for two weeks.

Talks and Seminars

Knowledge Graph Retrieval for Augmenting Language Models
ETRI Oct 2023
Direct Fact Retrieval from Knowledge Graphs & Personalized Subgraph Federated Learning
POSTECH ML Lab Jul 2023
Representation Learning on Graphs with Their Real-World Applications
Modulabs Jul 2022
Edge Representation Learning with Hypergraphs
KAIST AI Workshop Jan 2022
Accurate Learning of Graph Representations with Graph Multiset Pooling
KAIST AI Workshop Jan 2022
ICLR Social - ML in Korea May 2021
Introduction of Graph Neural Networks
Modulabs Jan 2022
Special Talk in Artificial Intelligence as an Alumnus
Korea University Aug 2021
Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction
Korean Conference on Computer Vision (KCCV) Aug 2021
NeurIPS Social - ML in Korea Dec 2020
Basic Algorithms of Machine Learning and Deep Learning with Python Code Challenges
Microsoft Coding Education Volunteer Group Seminar Jul 2018
Microsoft Student Partner Newcomer Tech Seminar May 2018
IaaS (Infrastructure as a Service) in Microsoft Azure
Microsoft Student Partner Newcomer Tech Seminar Apr 2018
Phishing Website Detection with Microsoft Azure Machine Learning Studio
KUSISWALL: Joint Seminar of Korea, Sookmyung and Hanyang Universities Feb 2018
Basic Algorithms of Machine Learning with Azure Machine Learning Studio
Microsoft Student Partner Evangelism Nov 2017
Microsoft Student Partner Evangelism Nov 2016
Azure Cognitive Services with Serverless Cloud Computing using JavaScript and Python
Microsoft Student Partner Evangelism Nov 2017
Windows Hello: Microsoft Face Recognition System in Windows 10
Microsoft Student Partner Inner Tech Seminar Oct 2016
Security Issues in Hash Encryption and Linux Password
KUSISWALL: Joint Seminar of Korea, Sookmyung and Hanyang Universities Jun 2016

Extra-curricular Activities

Lead & Representative of Korea, Microsoft Student Partner Jul 2016 - Jul 2020
- As a leader, promoted Microsoft products, recruited members, planned programs, and reviewed projects.
- As a representative of Korea on a global summit program, invited to Microsoft Build in 2017 which is the biggest annual conference held by Microsoft, and conducted a role as a technical ambassador.

Chairman, Inc0gnito Hacking Conference Mar 2018 - Feb 2019
- Organized the information security conference in Korea held by undergraduate students over 10 universities, which covers techs, news, and trends on information security areas.

8th Trainee, SW Maestro Jul 2017 - Dec 2017
- Researched and developed systems that automatically block obscene and violent movie content with machine learning, within the SW Maestro program that aims to train talented students on software in Korea.

Skills

Languages: Korean (mother tongue), English (fluent)
Programming (Coding): Python, JavaScript, Java, and C