Jinhyuk Lee

Ph.D. in Computer Science and Engineering
Visiting Postdoctoral Research Associate @ Princeton University
Research Professor @ Korea University

jinhyuk_lee@korea.ac.kr
`https://jhyuklee.github.io`
Last updated: 2021/08/31

RESEARCH INTEREST

My research area is based on natural language processing and deep learning. Specifically, I'm interested in learning **generalizable text retrieval with dense vectors** [ACL'19, ACL'21, EMNLP'21[(1)], EMNLP'21[(2)]] and tackling **challenges in biomedical NLP** [Bioinformatics'20, ACL'20, EMNLP'21[(3)]]. I'm also broadly interested in other NLP researches such as debiasing NLP models [EMNLP'20] and learning contextualized sparse representations [ACL'20].

EXPERIENCE

Princeton University
Visiting Postdoctoral Research Associate (PI: Danqi Chen)
Jan 2021 - Present
New Jersey, United States

· Leading projects on dense text retrieval [1, 4]
· Participating research on generalization of dense retrieval [2]

Korea University
Research Professor (PI: Jaewoo Kang)
Sep 2019 - Present
Seoul, Korea

· Leading a project on automatic knowledge base construction from question answering models [4]
· Advising research on QA, biomedical text mining, NLP applications [3, 5, 9]
· Teaching Data Science lecture (CVH126) and fulfilling mandatory military service

NAVER Clova
Research Intern (Host: Minjoon Seo)
Nov 2018 - Feb 2019, July 2019 - Sep 2019
Seongnam, Korea

· Co-leading a project on real-time open-domain question answering [10]
· Leading a project on contextualized sparse representations [8]
· Leading a project for pre-trained biomedical language model [2]

Samsung Electronics
Research Collaborator
Oct 2017 - May 2018
Seoul, Korea

· Co-leading a research project on anomaly detection in a semiconductor fab [6]
· Developing an anomaly detection system with multivariate time-series data using a deep neural network

Konolabs, Inc.
Research Intern
Mar 2016 - Aug 2016
Seoul, Korea

· Co-leading a research on event scheduling system with deep learning models [13]
· Developing a deep learning-based calendar scheduling system and android mobile application

Tagme, Inc.
Startup CEO
Aug 2014 - Feb 2016
Seoul, Korea

· Leading a startup company that provides cloud storage service with memorable file links
· Participating in the development of Web and mobile application of the product

EDUCATION

Korea University   Mar 2016 - Aug 2019
Ph.D. in Computer Science and Engineering (Advisor: Jaewoo Kang)
Thesis: Towards Domain-Specific Question Answering with Large Unstructured Knowledge Source

Korea University   Mar 2011 - Feb 2016
B.S. in Computer Science and Engineering

State University of New York Korea   June 2013 - Aug 2013
Software Technology and Enterprise (Summer School)

PUBLICATIONS

International Conferences

[1] **Phrase Retrieval Learns Passage Retrieval, Too**
Jinhyuk Lee, Alexander Wettig, Danqi Chen
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021 - long

[2] **Simple Entity-Centric Questions Challenge Dense Retrievers**
Christopher Sciavolino, Zexuan Zhong, **Jinhyuk Lee**, Danqi Chen
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021 - short

[3] **Can Language Models be Biomedical Knowledge Bases?**
Mujeen Sung, **Jinhyuk Lee**, Sean Yi, Minji Jeon, Sungdong Kim, Jaewoo Kang
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2021 - short

[4] **Learning Dense Representations of Phrases at Scale**
Jinhyuk Lee, Mujeen Sung, Jaewoo Kang, Danqi Chen
Annual Meeting of the Association for Computational Linguistics (ACL) 2021 - long (oral)

[5] **Look at the First Sentence: Position Bias in Question Answering**
Miyoung Ko, **Jinhyuk Lee**, Hyunjae Kim, Gangwoo Kim, Jaewoo Kang
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020 - long

[6] **Adversarial Subword Regularization for Robust Neural Machine Translation**
Jungsoo Park, Mujeen Sung, **Jinhyuk Lee**, Jaewoo Kang
Findings of Empirical Methods in Natural Language Processing (Findings of EMNLP) 2020 - short

[7] **Answering Questions on COVID-19 in Real-Time**
Jinhyuk Lee, Sean S Yi, Minbyul Jeong, Mujeen Sung, Wonjin Yoon, Yonghwa Choi, Miyoung Ko, Jaewoo Kang
Conference on Empirical Methods in Natural Language Processing (EMNLP), NLP COVID-19 2020 - long (oral)

[8] **Contextualized Sparse Representations for Real-Time Open-Domain Question Answering**
Jinhyuk Lee, Minjoon Seo, Hannaneh Hajishirzi, Jaewoo Kang
Annual Meeting of the Association for Computational Linguistics (ACL) 2020 - short (oral)

[9] **Biomedical Entity Representations with Synonym Marginalization**
Mujeen Sung, Hwisang Jeon, **Jinhyuk Lee**, Jaewoo Kang
Annual Meeting of the Association for Computational Linguistics (ACL) 2020 - long

[10] **Real-Time Open-Domain Question Answering on Wikipedia with Dense-Sparse Phrase Index**
Minjoon Seo[†], **Jinhyuk Lee**[†], Tom Kwiatkowski, Ankur Parikh, Ali Farhadi, Hannaneh Hajishirzi
Annual Meeting of the Association for Computational Linguistics (ACL) 2019 - long (oral)

[11] **Pre-trained Language Models for Biomedical Question Answering**
Wonjin Yoon, **Jinhyuk Lee**, Donghyeon Kim, Minbyul Jeong, Jaewoo Kang
Joint European Conference on Machine Learning and Knowledge Discovery in Databases, BioASQ 2019
Winner of the Seventh BioASQ Challenge - Task 7b, Phase B (1st place) [link]

[12] **Ranking Paragraphs for Improving Answer Recall in Open-Domain Question Answering**
Jinhyuk Lee, Seongjun Yun, Hyunjae Kim, Miyoung Ko, Jaewoo Kang
Conference on Empirical Methods in Natural Language Processing (EMNLP) 2018 - short (oral)

[13] **Learning User Preferences and Understanding Calendar Contexts for Event Scheduling**
Donghyeon Kim[†], **Jinhyuk Lee**[†], Donghee Choi, Jaehoon Choi, Jaewoo Kang
ACM International Conference on Information and Knowledge Management (CIKM) 2018 - full (oral)

[14] **Name Nationality Classification with Recurrent Neural Networks**
Jinhyuk Lee, Hyunjae Kim, Miyoung Ko, Donghee Choi, Jaehoon Choi, Jaewoo Kang
International Joint Conference on Artificial Intelligence (IJCAI) 2017 - full (oral)

International Journals

[1] **Crowdsourced identification of multi-target kinase inhibitors for RET- and TAU- based disease: the Multi-Targeting Drug DREAM Challenge**
Zhaoping Xiong, Minji Jeon, Robert J Allaway, Jaewoo Kang, Donghyeon Park, **Jinhyuk Lee** et al. PLOS Computational Biology 2021

[2] **BioBERT: a Pre-trained Biomedical Language Representation Model for Biomedical Text Mining**
Jinhyuk Lee[†], Wonjin Yoon[†], Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, Jaewoo Kang
Bioinformatics 2020
Best Paper for the Natural Language Processing Section of the 2020 IMIA Yearbook [link]

[3] **ReSimNet : Drug Response Similarity Prediction based on Siamese Neural Network**
Minji Jeon[†], Donghyeon Park[†], **Jinhyuk Lee**, Hwisang Jeon, Miyoung Ko, Sunkyu Kim, Yonghwa Choi et al.
Bioinformatics 2019
Winner of Multi-Targeting Drug DREAM Challenge (1st place) [link]

[4] **A Neural Named Entity Recognition and Multi-Type Normalization Tool for Biomedical Text Mining**
Donghyeon Kim, **Jinhyuk Lee**, Chan Ho So, Hwisang Jeon, Minbyul Jeong, Yonghwa Choi, Wonjin Yoon et al.
IEEE Access 2019

[5] **CollaboNet : Collaboration of Deep Neural Networks for Biomedical Named Entity Recognition**
Wonjin Yoon[†], Chan Ho So[†], **Jinhyuk Lee**, Jaewoo Kang
BMC Bioinformatics 2019

[6] **DeepNAP : Deep Neural Anomaly Pre-detection on a Semiconductor Fab**
Chunggyeom Kim[†], **Jinhyuk Lee**[†], Raehyun Kim, Youngbin Park, Jaewoo Kang
Information Sciences 2018

[†]: equal contributions

HONORS & AWARDS

Best Paper for the Natural Language Processing Section of the 2020 IMIA Yearbook [link]	2020
Winner of the Seventh BioASQ Challenge - Task 7b, Phase B (1st place) [link, press] · An international challenge on biomedical question answering hosted by Google, NIH, EU · Outperformed Google (2nd place), Peking University, Fudan University, and more	Sep 2019
Winner of Multi-Targeting Drug DREAM Challenge (1st place) [link, press] · An international challenge on multi-targeting drug hosted by IBM and Sage Bionetworks · Outperformed Janssen Pharmaceutica (4th place), Bayer, and more	Nov 2018
Korea University Best Honors Scholarships	Mar 2014
The Grand Prize, Start-up Competition, Korea University	Jan 2014
Venture Internship Program Scholarships, Korea University	Dec 2013

PATENTS

Method and System for Discovery New Drug Candidate
Jaewoo Kang, Minji Jeon, Donghyeon Park, **Jinhyuk Lee**, Hwisang Jeon, Miyoung Ko
10-2019-0053667

Device and Method for Providing Nationality Information of User Name using Neural Networks
Jaewoo Kang, Jaehoon Choi, **Jinhyuk Lee**, Donghee Choi
10-2017-0147399

Method, system and non-transitory computer-readable recording medium for providing schedule-related advertisements
Yun Jeong Min, Jaehoon Choi, Donghyeon Kim, **Jinhyuk Lee**
10-2016-0173499

Method, system and non-transitory computer-readable recording medium for providing predictions on calendar
Jaewoo Kang, Donghyeon Kim, **Jinhyuk Lee**, Jaehoon Choi
10-2016-0173483

SERVICES

Conference/Workshop Reviewer
ACL 2020-2021, **EMNLP 2020-2021**, **NAACL 2021**, **AKBC 2020-2021**, **2nd MRQA (EMNLP 2019)**

Journal Reviewer
PLOS ONE, **IEEE Access**, **IEEE TCBB**

INVITED TALKS

Learning Dense Representations of Phrases at Scale
co:here - Aug 2021
Apple - May 2021
Google - Feb 2021

Biomedical Natural Language Processing
Inha University - May 2021
APTOS 2020: Natural Language Processing in Medicine - Dec 2020
Korean AI Association 2020 - Nov 2020

Open-Domain Question Answering with Phrase Retrieval
Korea-Japan Database Workshop 2019 @ Tokyo Tech - Oct 2019

TEACHING

CVH126 Introduction to Data Science, Korea University   Fall 2019
Course Lecturer

COSE214 Algorithms, Korea University   Fall 2016
Teaching Assistant

REFERENCES

Available on request.