

Dongyeop Kang

☎ +1-412-736-4873 | ✉ dongyeopk@berkeley.edu | 🏠 <https://dykang.github.io/>

EDUCATION

Carnegie Mellon University PH.D, Language Technology Institute, School of Computer Science	<i>Pittsburgh, PA, USA</i> 2011, 2015 - 2020
▪ Advisor: Eduard Hovy ▪ Thesis: "Linguistically Informed Language Generation: A Multifaceted Approach" ▪ Leave-of-absence (2012-2015) for alternative military service in S. Korea	
Korea Advanced Institute of Science and Technology M.S. / B.S., Computer Science Engineering	<i>Daejeon, Korea</i> 2010
▪ Thesis: "Multidimensional Mining of Search Logs based on Hierarchical Latent Variable Modeling" ▪ Summa Cum Laude, 1st place in CSE undergraduate students	
Seoul Science High School	<i>Seoul, Korea</i>

PROFESSIONAL APPOINTMENTS

University of California, Berkeley Postdoctoral Scholar PI: Marti A. Hearst	<i>Berkeley, CA, USA</i> 2020 - Present
Carnegie Mellon University Research Assistant Language Technology Institute Advisor: Eduard Hovy	<i>Pittsburgh, PA, USA</i> 2015 - 2020
Facebook AI Research Research Intern Mentors: Y-Lan Boureau, Jason Weston	<i>Menlo Park, CA, USA</i> May 2018 - Aug 2018
Allen Institute for Artificial Intelligence (AI2) Research Intern Mentors: Tushar Khot, Ashish Sabharwal, Peter Clark	<i>Seattle, WA, USA</i> Jul 2017 - Sep 2017
Microsoft Research Research Intern KTG Group Mentors: Michael Gamon, Patrick Pantel	<i>Redmond, WA, USA</i> Jun 2016 - Aug 2016
Naver Labs Assistant Manager Data Science Group (Alternative Military Service)	<i>Seongnam, Korea</i> 2014 - 2015
KAIST Institute Researcher knowledge convergence team (Alternative Military Service)	<i>Daejeon, Korea</i> 2012 - 2014
Oracle Labs Research Assistant Project-Q group Mentor: Sungpack Hong	<i>Redwood City, CA, USA</i> May 2014 - Aug. 2014
IBM T. J. Watson Research Center Research Intern Information Security Group Mentor: Youngja Park	<i>Yorktown, NY, USA</i> Jun - Oct 2013
Eventrader Co-founder funded by KAIST Institute and SMBA	<i>Seoul, Korea</i> 2013 - 2015
Hong Kong University of Science and Technology Research Assistant Computer Science Engineering Dept. Mentor: Sung Kim	<i>Hong Kong</i> Feb 2011 - Aug 2011
Microsoft Research Research Intern Operating Systems Group Mentor: Edmund B Nightingale	<i>Redmond, WA, USA</i> Jun - Sep 2010
Microsoft Research Asia Research Assistant IR and Mining Group Mentors: Daxin Jiang, Hang Li	<i>Beijing, China</i> Aug 2009 - May 2010
Korea Advanced Institute of Science and Technology Undergrad Research Assistant Computer Science Dept. Advisor: Ho-Jin Choi	<i>Daejeon, Korea</i> 2005 - 2009

HONORS & AWARDS

2018	Top Reviewers, NeurIPS 2018 and ACL 2018	
2018	Allen Institute for AI Research Fellowship, Full tuition + stipend	<i>Seattle, WA, USA</i>

2016-2017	CMU Presidential Fellowship, Full tuition + stipend	Pittsburgh, PA, USA
2011-2016	ILJU Foundation Scholarship, Ph.D Fellowship	Seoul, Korea
2013	SMBA Startup Investment Prize, 100k USD for one year of support	Seoul, Korea
2010	Excellent Performance Award, Microsoft Research Asia	Beijing, China
2007	Summa Cum Laude, 1st place in CSE dept. KAIST	Daejeon, Korea

PUBLICATIONS

(* means equal contribution)

- Submitted Does BERT know this word's definition? A Comparative Study of Contextual Embedding Representations and Formal Dictionary Definitions, Hwiyeol Jo, Dongyeop Kang, Byoung-Tak Zhang, Marti Hearst, Submitted to EACL 2021
- Submitted Unsupervised Natural Language Video Localization, Jinwoo Nam, Daechul Ahn, Dongyeop Kang, Seong Jong Ha, Jonghyun Choi, Submitted to CVPR 2021
- CHI'21 Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols, Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, Marti A. Hearst, SIGCHI Conference on Human Factors in Computing Systems (CHI), 2021
- DeeLIO'20 GenAug: Data Augmentation for Finetuning Text Generators, Steven Y. Feng*, Varun Gangal*, Dongyeop Kang, Teruko Mitamura, Eduard Hovy, Deep Learning Inside Out (DeeLIO) Workshop at EMNLP 2020
- SDP'20 Document-Level Definition Detection in Scholarly Documents: Existing Models, Error Analyses,, Future Directions, Dongyeop Kang Andrew Head, Risham Sidhu, Kyle Lo, Daniel Weld, Marti A. Hearst, Workshop on Scholarly Document Processing (SDP) at EMNLP 2020
- EMNLP'20 Plan ahead: Self-Supervised Text Planning for Paragraph Completion Task, Dongyeop Kang, Eduard Hovy, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020
- EMNLP'20 INSPIRED: Toward Sociable Recommendation Dialog Systems, Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, Zhou Yu, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020
- Thesis Linguistically Informed Language Generation: A Multifaceted Approach, Dongyeop Kang, Ph.D. Dissertation, Carnegie Mellon University (Committee: Eduard Hovy, Jeffrey Bigham, Alan W Black, Jason Weston, Dan Jurafsky)
- ACL'20 Posterior Calibrated Training on Sentence Classification Tasks, Taehee Jung, Dongyeop Kang, Hua Cheng, Lucas Mentch, Thomas Schaaf, Annual Meeting of the Association for Computational Linguistics (ACL), 2020
- Preprint xSLUE: A Benchmark, Analysis Platform for Cross-Style Language Understanding and Evaluation, Dongyeop Kang, Eduard Hovy, arXiv 1911.03663, Submitted to EACL 2021
- Preprint Don't Treat Every Relation the Same! Constraining Word Vectors to Geometric Properties of Inter-Word Relations, Dongyeop Kang*, Taehee Jung*, Eduard Hovy, 2019
- EMNLP'19 Recommendation as a Communication Game: Self-Supervised Bot-Play for Goal-oriented Dialogue, Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, Jason Weston, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019
- EMNLP'19 (Male, Bachelor), (Female, Ph.D) have different connotations: Parallely Annotated Stylistic Language Dataset with Multiple Personas, Dongyeop Kang, Varun Gangal, Eduard Hovy, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019
- EMNLP'19 Linguistic Versus Latent Relations for Modeling Coherent Flow in Paragraphs, Dongyeop Kang, Hiroaki Hayashi, Alan W Black, Eduard Hovy, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019
- EMNLP'19 Earlier Isn't Always Better: Sub-aspect Analysis on Corpus, System Biases in Summarization, Dongyeop Kang*, Taehee Jung*, Lucas Mentch, Ed Hovy (*equal contribution), Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019
- EMNLP'18 Bridging Knowledge Gaps in Neural Entailment via Symbolic Models, Dongyeop Kang, Tushar Khot, Ashish Sabharwal, Peter Clark, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2018

- ACL'18 AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples, Dongyeop Kang, Tushar Khot, Ashish Sabharwal, Eduard Hovy, Annual Meeting of the Association for Computational Linguistics (ACL), 2018
- NAACL'18 A Dataset of Peer Reviews (PeerRead): Collection, Insights, NLP Applications, Dongyeop Kang, Waleed Ammar, Bhavana Dalvi, Roy Schwartz, Eduard Hovy, Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL), 2018
- AAAI'18 Actionable email intent modeling with reparametrized RNN, Chu-Cheng Lin, Dongyeop Kang, Michael Gamon, Patrick Pantel, Association for the Advancement of Artificial Intelligence (AAAI), 2018
- EMNLP'17 Detecting, Explaining Causes From Text For a Time Series Event, Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, Eduard Hovy, Conference on Empirical Methods in Natural Language Processing (EMNLP), 2017
- RecSysW'15 News2Images: Automatically Summarizing News Articles into Image-Based Contents via Deep Learning, Jung-Woo Ha, Dongyeop Kang, Hyuna Pyo, Jeonghee Kim, Workshop on News Recommendation, Analytics in RecSys, 2015
- ICDM'14 Eventera: Real-time Event Recommendation System from Massive Heterogeneous Online Media, Dongyeop Kang, Donggyun Han, Na Hea Park, Sangtae Kim, U Kang, Soobin Lee, IEEE International Conference on Data Mining (ICDM), 2014
- CIKM'14 Data/Feature Distributed Stochastic Coordinate Descent for Logistic Regression, Dongyeop Kang, Woosang Lim, Kijung Shin, Sael Lee, U Kang, ACM International Conference on Information, Knowledge Management (CIKM), 2014
- ECML'14 Hetero-Labeled LDA: A partially supervised topic model with heterogeneous label information, Dongyeop Kang, Youngja Park, Suresh Chari, European Conference on Machine Learning, Principles (ECML), 2014
- WSDM'11 Multidimensional Mining of Large-Scale Search Logs: A Topic-Concept Cube Approach, Dongyeop Kang, Daxin Jiang, Jian Pei, Zhen Liao, Xiaohui Sun, HJ Choi, ACM International Conference on Web Search, Data Mining (WSDM), 2011

TEACHING AND TALKS

- 2020 Nov KAIST AI Colloquium, Talk about "Humans in Language Generation"
- 2020 Nov GIST EECS Colloquium, Talk about "Toward Human-like Language Generation"
- 2020 Jun Berkeley AI Research (BAIR) Workshop, Talk about "Uncertainty Calibration in NLP Classifiers"
- 2020 May CMU LTI Summer Seminar, Talk about "Multi-faceted Approach for Language Generation"
- 2020 May POSTECH AI Seminar, Talk about "Toward Human-like Language Generation"
- 2018 Spring Computational Semantics for NLP, teaching assistant
- 2017 Spring Machine Translation and Sequence to Sequence Models, teaching assistant
- 2010 Spring Artificial Intelligence, teaching assistant

ACADEMIC SERVICES

- Program ACL (2018-2020), EMNLP (2018-2020), NAACL (2019), NeurIPS (2018-2020), ICLR (2019-2021), ICML
Committee (2019), MRQA (2018), WNUT (2019), Scientometrics (2019),

REFEREES

Eduard Hovy
Professor, CMU / DARPA
hovy@cmu.edu
310-529-3967

Marti A. Hearst
Professor, UC Berkeley
hearst@berkeley.edu
510-642-8016

Alan Black
Professor, CMU
awb@cs.cmu.edu
412-268-6299

Patrick Pantel
Director, Facebook AI
ppantel@fb.com
425-882-8080