Don (Dong Won) Lee

dongwonl.com | Phone: 747-258-7999 | E-Mail: dongwonl@cs.cmu.edu

Education

Carnegie Mellon University   **Pittsburgh, PA**
M.S. Machine Learning   Expected Grad.: May, 2022

Carnegie Mellon University   **Pittsburgh, PA**
B.S. Statistics and Machine Learning   May, 2021

Relevant Coursework: Introduction. to ML (Ph.D.), Deep Reinforcement Learning, Multimodal Machine Learning, Convex Optimization, Machine Learning in Practice, Probability and Mathematical Statistics, Probabilistic Graphical Models, Advanced Multimodal Machine Learning, Scalability in Machine Learning

Publications

Published

- **D. Lee**, C. Ahuja, R. Ishii, and L. Morency, "Crossmodal clustered contrastive learning: Grounding of spoken language to gestures"   (ICMI GENEA Workshop, 2021)
- C. Ahuja, **D. Lee**, Y. Nakano, and L. Morency, "Style Transfer for Co-Speech Gesture Animation: A Multi-Speaker Conditional-Mixture Approach"   (ECCV 2020)
- C. Ahuja, **D. Lee**, R. Ishii, and L. Morency, "No Gestures Left Behind: Learning Relationships between Spoken Language and Freeform Gestures"   (Findings EMNLP 2020)

In Submission / Preparation

- C. Ahuja, **D. Lee,** and L. Morency, "Low-Resource Adaptation of Spatio-Temporal Crossmodal Generative Models"   (In Submission at CVPR 2022)
- **D. Lee**, I. Grover, P. Colon-Hernandez, HW. Park, C. Breazeal, "Role-Aware Graphical Next Speaker Prediction in Multimodal Multiparty Interaction"   (Rejected at HRI 2022, in revision for ACM CSCW 2022)
- **D. Lee**, C. Ahuja, P. Liang**,** and L. Morency, "Localized Presentations Dataset: Exploring Multimodality in Lectures"   (In Preparation for NAACL 2022)
- **D. Lee**, B. Eysenbach, C. Lynch, S. Levine, R. Salakhutinov, "Language-Conditioned Control via Recursive Classification of Natural Language"   (In Preparation for ICML 2022)

Research

Language Technologies Institute, Carnegie Mellon University   **Pittsburgh, PA**
Generation of Human Poses Conditioned on Language (Prof. L.P Morency)   Jun. 2019 ~ Present

- Wrote a first-author paper and 3 second-author papers on generating human co-speech gestures which are well-timed and relevant to the speaker's spoken language
- Implemented a multimodal transformer after considering varying scales of positional encodings (word-level, frame-level), this model outperformed previous state-of-the-art approaches in coverage metric (F1) by 5 percent
- Proposed a new cross-modal contrastive learning approach to enforce many-to-one grounding between spoken language and gestures, which outperformed state-of-the-art models in precision (L1) by 11 percent and presented findings at the 2nd GENEA Workshop at ICMI 2021
- Collected and released a large human gestures dataset, PATS (Pose, Audio, Transcript, Style), which includes 25 speakers and 251 hours of data to study the relationship between co-speech gestures with audio and text signals
- Wrote code to download and preprocess the data into the required format, from raw video to transcriptions (ASR) then to BERT (language), log-mel-spectrograms (audio), skeletal keypoints (pose)
- Organized the 1st Workshop on Crossmodal Social Animation at ICCV 2021 as the publication chair, where I reviewed submitted papers and jointly presided over the workshop

Lecture Video Captioning (Prof. L.P Morency) Jan. 2021 ~ Present
- Collected a new dataset of educational videos with 197 hours of data, 336 videos, a total number of 22,000 slides to build a machine learning model can explain slides with natural language.
- Curated data for mouse traces to be used as grounding signals, as we cannot rely on annotations from a crowd-sourcing service due to required expert-level knowledge.
- Preprocessed data by extracting the layout (figure, text), spoken language to text (ASR), mouse traces of lecturer to provide grounding signal (optical flow)
- Explored various levels of input data (slide level, and figure level) and utilized Multi-Instance Learning to address the weak alignment between captions and lecture slides for the task of slide to image and image to slide retrieval, improving recall metrics by 3 percent
- Re-implemented state-of-the-art captioning model MITR to extract baseline metrics for lecture slide captioning

Personal Robots Group, Massachusetts Institute of Technology **(Remote) Boston, MA**
Human-Robot Engagement Prediction (Dr. Hae Won Park, Prof. Cynthia Breazeal) May 2021 ~ Present
- Built a novel graphical model of a conversational group to improve performance on next speaker identification by 8 percent in accuracy, 11 percent in F1 scores, which is planned for submission at CSCW 2022
- Developed entire codebase, implemented model and experimental set-up, wrote first-author paper and submitted rebuttal at HRI 2022 (rejected)
- Utilized findings from Psychology that there are behavioral differences for each role in the group, and conditioned the representations on each speaker's role
- Included both nonverbal and verbal features for next speaker prediction and ran ablations to understand which feature which feature affects predictions

Machine Learning Department, Carnegie Mellon University **Pittsburgh, PA**
Language Guided Goal Conditioned RL (Prof. Ruslan Salakhutinov) Jan. 2021 ~ Present
- Implemented a novel reinforcement learning algorithm which takes in a desired natural language description of a future state to infer the necessary actions to reach the goal state via recursive classification
- Devised new sampling strategy (sampling state-action pairs in the time steps prior to the desired action) to prevent naive success cases to be exposed frequently
- Developed the entire codebase for 3 RL environments, implemented model and experimental set-up
- Solved the task of embodied pick, navigation, manipulation (Mujoco) with proposed approach

Teaching

Machine Learning Department, Carnegie Mellon University **Pittsburgh, PA**
Convex Optimization (Taught by Prof. Yuanzhi Li) (1 Semester) Feb. 2021 ~ May 2021
- Served as a TA for over 150 students for a PhD level course as an undergraduate student
- Led weekly office hours to clarify advanced mathematical concepts and theory, graded proofs and code

Statistics and Data Science Department, Carnegie Mellon University **Pittsburgh, PA**
Statistics & Data Science Methods (Taught by Prof. Gordon Weinberg) (4 Semesters) Aug. 2019 ~ May 2021
- Taught 2 classes: Statistical Reasoning and Practice and Statistical Methods
- Led weekly R lab sessions where students practiced statistical methods, hosted weekly office hours to clarify concepts, graded exams and homework for over 200 students in each class

Experience

Special Operations Command, Republic of Korea Army **Abu Dhabi, UAE**
Deployed Special Operations Unit Interpreter Jun. 2017 ~ May. 2019
- Selected (2 out of 192 candidates) based on work performance and physical tests to serve as interpreter for AKH 14 Korean Special Operations Unit UAE
- Led and organized combined mountain warfare training for UAE Special Operations soldiers and daily situation briefs and after-action reviews
- Received Certificate of Commendation (1), Certificate of Excellence (2)

Professional Services & Awards

Workshop Publication Chair: 1st Workshop on Crossmodal Social Animation (ICCV 2021)
Reviewer: EMNLP 2021, ICMI 2021
Awards: Undergraduate Research Grant (CMU, 2019), Certificate of Commendation, Excellence (ROK Spec. Ops)

Skills

Programming Languages: (Well-Experienced in) Python, R, Javascript, HTML, CSS
Machine Learning Frameworks: (Well-Experienced in) PyTorch, TensorFlow, Scikit-Learn, Keras, Pandas, Numpy
Language: (Native) English, Korean, (Fluent) Mandarin