SpLU-RoboNLP 2024

The 4th Workshop on Spatial Language Understanding and Grounded Communication for Robotics

Proceedings of the Workshop

August 11-16, 2024

Order copies of this and other ACL proceedings from:

Introduction

Leveraging the foundation built in the prior workshops SPLU-RoboNLP-2022, SPLU-RoboNLP-2021, SpLU 2020, SpLU-RoboNLP 2019, SpLU 2018, and RoboNLP 2017, we organize the fourth combined workshop on Spatial Language Understanding and Grounded Communication for Robotics, SpLU-RoboNLP-2024. To achieve the long-term goal of natural conversation with robots in our homes, workplaces, hospitals, and warehouses, it is essential that we develop new techniques for linking language to perception and actions in the physical world.

This requires developing tools and theories to find insights into addressing some fundamental questions in NLP and HRI. Some important questions are the following. Can we give instructions to robotic agents to assist with navigation and manipulation tasks in remote settings? Can we talk to robots about the surrounding physical world, and help them interactively learn the language needed to finish a task? Can we develop robots that reply to us via grounded language generation, and eventually lead to an effective, two-way grounded dialogue? Given the rise of generative large language models, another question is how these large models can be deployed in situated dialogue settings and act meaningfully.

Human-robot dialogue often involves developing an understanding of grounded spatial descriptions. These capabilities invariably require understanding spatial semantics that relate to the physical environments where robots are embodied. Spatial semantics are the part of language semantics that is most related to grounding language into perception and the physical world. Spatial language meaning representation includes research related to cognitive and linguistically motivated spatial semantic representations, spatial knowledge representation and ontologies, qualitative and quantitative representation models, spatial annotation schemes, and efforts for creating specialized corpora. Spatial language learning considers both symbolic and sub-symbolic (with continuous representations) techniques and computational models for spatial information extraction, semantic parsing, and spatial co-reference within a global context that includes discourse and pragmatics from data or formal models. Recent studies show that one of the semantic aspects that pre-trained language models and even the recent large generative language models struggle with is reasoning over spatial language. We are interested in investigating whether qualitative and quantitative formal representations are helping spatial reasoning based on natural language and the possibility of learning such representations from data. Moreover, we emphasize on multimodality aspect of spatial language understanding as well as human-robot interaction. Some interesting related questions include, *which representations are appropriate for different modalities, and which ones are modality independent? How can we exploit visual information for language learning and reasoning?* The main goal of this joint workshop is to bring in the perspectives of researchers working on physical robot systems and with human users and align spatial language understanding representation and learning approaches, datasets and benchmarks with the goals and constraints encountered in HRI and robotics. Such constraints include high costs of real-robot experiments, computational costs for real-time interactions, human-in-the-loop training and evaluation settings, scarcity of embodied data, as well as non-verbal communication.

The invited speakers, program committee, and organizing committee consist of researchers who belong to language, robotics, and vision communities or work in the intersection of these research areas.

We have 4 invited speakers, 3 archived papers, and several non-archival papers. Our workshop will accommodate the relevant ACL findings papers.

Organizing Committee

Organizing Committee

Parisa Kordjamshidi, Michigan State University
Xin Wang, University of California Santa Cruz
Yue Zhang, Michigan State University
Ziqiao Ma, University of Michigan
Mert Inan, Northeastern University

Program Committee

Program Committee

Table of Contents

Program

09:00 - 09:15   *Opening Remarks*

09:15 - 10:00   *Invited Talk1*

10:00 - 10:30   *Grounded Communication for Robotics*

         Language-guided World Models: A Model-based Approach to AI Control
Alex L Zhang, Khanh Xuan Nguyen, Jens Tuyls, Albert Lin and Karthik R Narasimhan

         Natural Language Can Facilitate Sim2Real Transfer
Albert Yu, Adeline Foote, Ray Mooney and Roberto Martín-Martín

10:30 - 11:00   *Coffee Break*

11:00 - 11:45   *Invited Talk2*

11:45 - 12:30   *Grounded Communication for Robotics*

         Into the Unknown: Generating Geospatial Descriptions for New Environments
Tzuf Paz-Argaman, John Palowitch, Sayali Kulkarni, Reut Tsarfaty and Jason Michael Baldridge

         Tuning Language Models with Spatial Logic for Complex Reasoning
Tanawan Premsri and Parisa Kordjamshidi

         TopViewRS: Vision-Language Models as Top-View Spatial Reasoners
Chengzu Li, Caiqi Zhang, Han Zhou, Nigel Collier, Anna Korhonen and Ivan Vulić

12:30 - 14:00   *Lunch*

14:00 - 14:45   *Invited Talk3*

14:45 - 15:30   *Invited Talk4*

Friday, August 16, 2024 (continued)

15:30 - 16:00   *Coffee Break*

16:00 - 16:20   *Poster Spotlight*

16:20 - 17:30   *Posters*