SpLU-RoboNLP 2021

The 2nd International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics

Proceedings of the Workshop

August 5-6, 2021
Bangkok, Thailand (online)

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-78-7

Introduction

SpLU-RoboNLP 2021 is a combined workshop on spatial language understanding (SpLU) and grounded communication for robotics (RoboNLP) that aims to realize the long-term goal of natural conversation with machines in our homes, workplaces, hospitals, and warehouses by highlighting developments in linking language to perception and actions in the physical world. It also highlights the importance of spatial semantics when it comes to communicating about the physical world and grounding language in perception. The combined workshop aims to bring together members of NLP, robotics, vision and related communities in order to initiate discussions across fields dealing with spatial language understanding and grounding language to perception and actions in the real world. 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, human-in-the-loop training and evaluation settings, scarcity of embodied data, as well as non-verbal communication.

Recent years have seen an increase in the availability of simulators in which virtual agents can take actions and obtain realistic visual observations, which has led to the creating of benchmarks for grounded language understanding in such environments. These benchmarks allow more direct comparisons of different techniques on certain tasks and have led to a significant increase in interest in some tasks such as vision and language navigation. However, many challenges still remain. Most systems using such benchmarks do not actually perform interactive training - obtaining live feedback from the environment on taking novel actions. Such training becomes more expensive as the simulator starts to support more actions. Different simulators and benchmarks vary in the extent to which they model realistic tasks or realistic capabilities of physical robots. Many of the modeling techniques used on such benchmarks may require too much compute to be used on physical robots.

Following the exciting recent progress in a number of visual language grounding tasks and vision and language navigation, the creation of more interactive embodied agents that can reason about spatial knowledge, common sense knowledge and information provided in instructions, generalize to data beyond what is seen during training, identify gaps in their knowledge or understanding, and engage in natural language interactions with users to fill in these gaps and explain their behavior are interesting research directions.

We have accepted 6 archival submissions and the workshop included an additional 4 non archival submissions.

Organizers:

Malihe Alikhani, University of Pittsburgh
Valts Blukis, Cornell University
Parisa Kordjamshidi, Michigan State University
Aishwarya Padmakumar, Amazon Alexa AI
Hao Tan, University of North Carolina, Chapel Hill

Program Committee:

Jacob Arkin, University of Rochester
Jonathan Berant, Tel-Aviv University
Steven Bethard, University of Arizona
Johan Bos, University of Groningen
Volkan Cirik, Carnegie Mellon University
Guillem Collell, KU Leuven
Simon Dobnik, University of Gothenburg, Sweden
Fethiye Irmak Dogan, KTH Royal Institute of Technology
Frank Ferraro, University of Maryland, Baltimore County
Daniel Fried, University of California, Berkeley
Felix Gervits, Tufts University
Yicong Hong, Australian National University
Drew Arad Hudson, Stanford University
Xavier Hinaut, INRIA
Gabriel Ilharco, University of Washington
Siddharth Karamcheti, Stanford University
Hyounghun Kim, University of North Carolina, Chapel Hill
Jacob Krantz, Oregon State University
Stephanie Lukin, Army Research Laboratory
Lei Li, ByteDance AI Lab
Roshanak Mirzaee, Michigan State University
Ray Mooney, University of Texas, Austin
Mari Broman Olsen, Microsoft
Natalie Parde, University of Illinois, Chicago
Christopher Paxton, NVIDIA
Roma Patel, Brown University
Nisha Pillai, University of Maryland, Baltimore County
Preeti Ramaraj, University of Michigan
Kirk Roberts, University of Texas, Houston
Anna Rohrbach, University of California, Berkeley
Mohit Shridhar, University of Washington
Ayush Shrivastava, Georgia Tech
Jivko Sinapov, Tufts University
Kristin Stock, Massey University of New Zealand
Alane Suhr, Cornell University
Rosario Scalise, University of Washington
Morgan Ulinski, Columbia University
Xin Wang, University of California, Santa Cruz
Shiqi Zhang, SUNY Binghamton

Invited Speakers:

Maja Matarić, University of Southern California

Kartik Narasimhan, Princeton University
Jean Oh, Carnegie Mellon University
Thora Tenbrink, Bangor University

Table of Contents

Conference Program

Friday, Aug 6, 2021 (EDT)

09:00 - 10:00 **Poster Session**

10:00 - 12:00 **Morning Invited Talks**

10:00 - 11:00 *Invited Talk*
Thora Tenbrink

11:00 - 12:00 *Invited Talk*
Jean Oh

12:00 - 13:50 **Morning Session**

Symbol Grounding and Task Learning from Imperfect Corrections
Mattias Appelgren and Alex Lascarides

Learning to Read Maps: Understanding Natural Language Instructions from Unseen Maps
Miltiadis Marios Katsakioris, Ioannis Konstas, Pierre Yves Mignotte and Helen Hastie

Visually Grounded Follow-up Questions: a Dataset of Spatial Questions Which Require Dialogue History
Tianai Dong, Alberto Testoni, Luciana Benotti and Raffaella Bernardi

Modeling Semantics and Pragmatics of Spatial Prepositions via Hierarchical Common-Sense Primitives
Georgiy Platonov, Yifei Yang, Haoyu Wu, Jonathan Waxman, Marcus Hill and Lenhart Schubert

Towards Navigation by Reasoning over Spatial Configurations
Yue Zhang, Quan Guo and Parisa Kordjamshidi

Learning to Parse Sentences with Cross-Situational Learning using Different Word Embeddings Towards Robot Grounding
Subba Reddy Oota, Frederic Alexandre and Xavier Hinaut

Error-Aware Interactive Semantic Parsing of OpenStreetMap
Michael Staniek and Stefan Riezler

Compositional Data and Task Augmentation for Instruction Following
Soham Dan, Xinran Han and Dan Roth

14:00 - 15:00 **ACL Findings Papers**

Language-Mediated, Object-Centric Representation Learning
Ruocheng Wang, Jiayuan Mao, Samuel Gershman, Jiajun Wu

Probing Image-Language Transformers for Verb Understanding
Lisa Anne Hendricks, Aida Nematzadeh

Hierarchical Task Learning from Language Instructions with Unified Transformers and Self-Monitoring
Yichi Zhang, Joyce Chai

VLM: Task-agnostic Video-Language Model Pre-training for Video Understanding
Hu Xu, Gargi Ghosh, Po-Yao Huang, Prahal Arora, Masoumeh Aminzadeh, Christoph Feichtenhofer, Florian Metze, Luke Zettlemoyer

Grounding 'Grounding' in NLP
Khyathi Raghavi Chandu, Yonatan Bisk, Alan W Black

PROST: Physical Reasoning of Objects through Space and Time
Stéphane Aroca-Ouellette, Cory Paik, Alessandro Roncone, Katharina Kann

15:00 - 16:00 **Panel Session**

17:00 - 19:00 **Afternoon Invited Talks**

17:00 - 18:00 *Invited Talk*
Karthik Narasimhan

18:00 - 19:00 *Invited Talk*
Maja Mataric

19:00 - 20:15 **Afternoon Session**

Plan Explanations that Exploit a Cognitive Spatial Model
Raj Korpan and Susan L. Epstein

Fine-Grained Spatial Information Extraction in Radiology as Two-turn Question Answering
Surabhi Datta and Kirk Roberts

Interactive Reinforcement Learning for Table Balancing Robot
Haein Jeon, Yewon Kim and Bo-Yeong Kang

Multi-Level Gazetteer-Free Geocoding
Sayali Kulkarni, Shailee Jain, Mohammad Javad Hosseini, Jason Baldridge, Eugene Ie and Li Zhang

Interactive learning from activity description
Khanh Nguyen, Dipendra Misra, Robert Schapire, Miro Dudík and Patrick Shafto

20:15 - 21:00 **Poster Session**