Semi-Supervised Deep Learning for Domain Adaptation in Robotic Language Acquisition

Cynthia Matuszek, Francis Ferraro
cmat, ferraro @ umbc.edu
University of Maryland, Baltimore County
iral.cs.umbc.edu

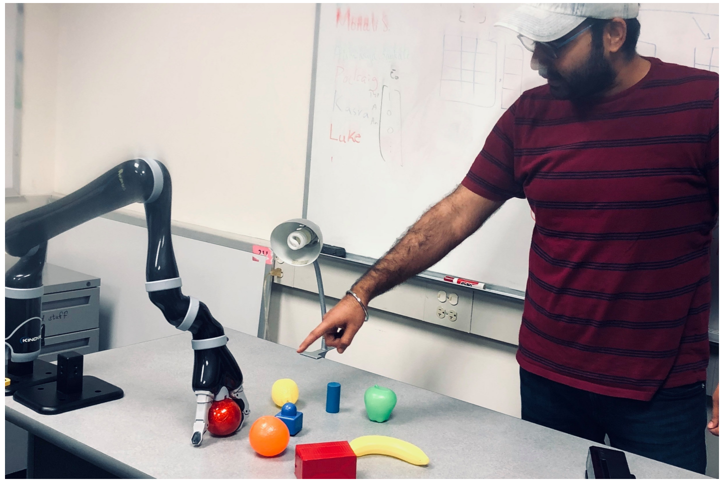
UMBC

2021 NRI & FRR Principal Investigators' Meeting
March 10-12, 2021

NSF #2024878, Oct. 2020–Oct 2023, UMBC

Project goals

- Collaboration requires communication
  - Natural language for HRI: intuitive and adaptable
- Teach about/instruct environment and teammates
  - Language in or about the world = grounded language
- Don't teach every robot, every time



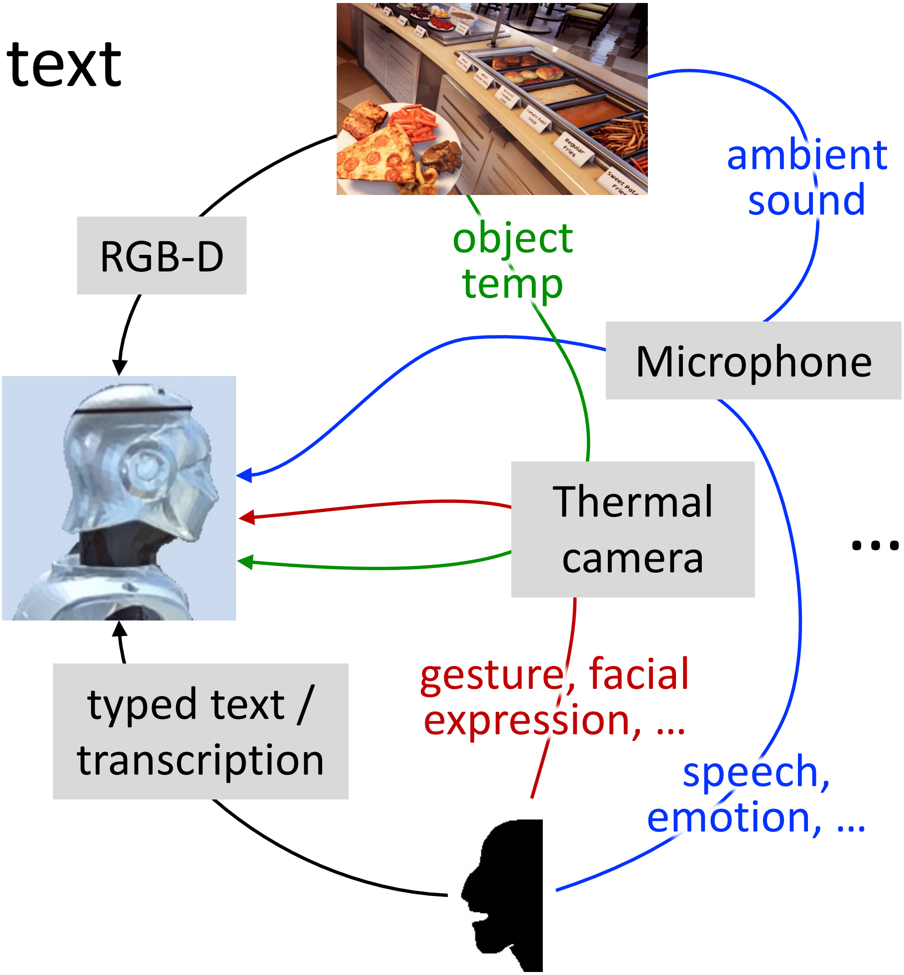
Goals:

1. Robots learn, from language, to perform tasks in human environments
2. Transfer this learned knowledge across platforms and tasks

NSF #2024878, UMBC, Cynthia Matuszek — cmat@umbc.edu

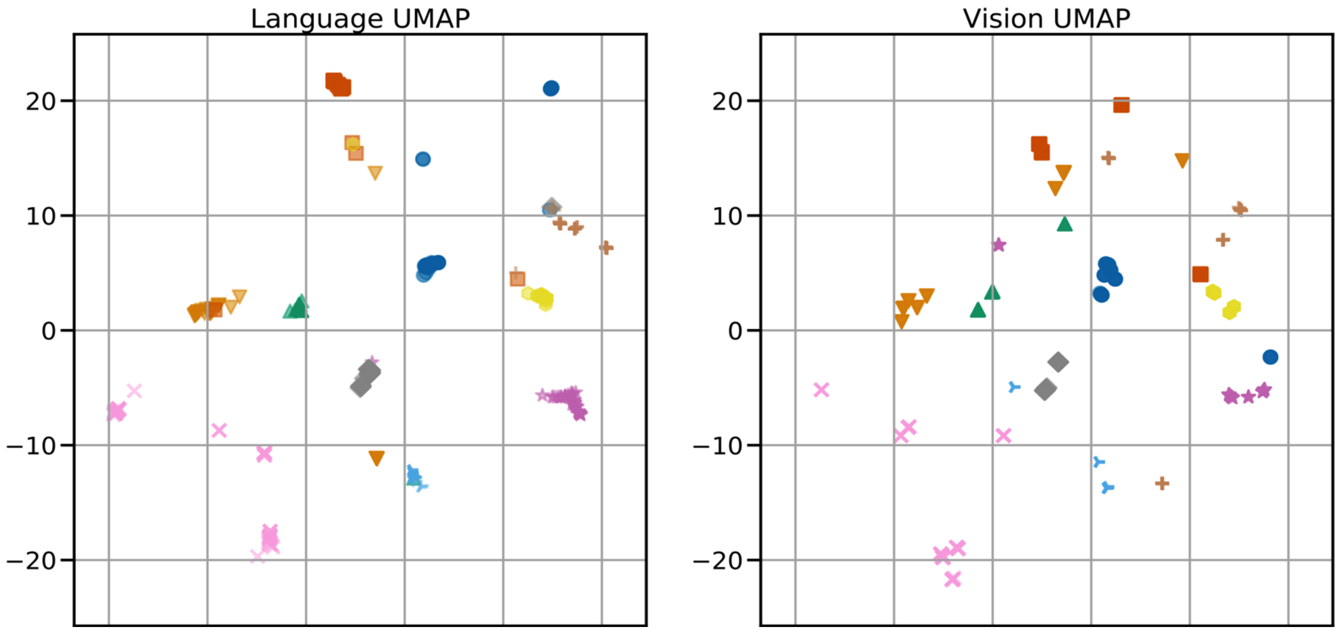
Learning language from multimodal percepts

- Current grounded language learning: sensors + text
  - RGB-D; infrequently, haptics, sound
  - Language: typed or transcribed speech (often constrained)
- Language is contextual
  - Human: gesture, gaze, body pose, …
  - Environmental factors: temperature, sound, …
  - Some have been incorporated individually
- Use multimodal sensing to directly link modalities to/from linguistic constructs



Learning from multimodal percepts (2)

- Direct geometric methods + dimensional affinity discovery
- Learn alignments among heterogeneous data streams
  - Learn alignments between language and sensor inputs
  - Pivot across language to find correspondences
- Induce a joint model of language and mixed sensor data

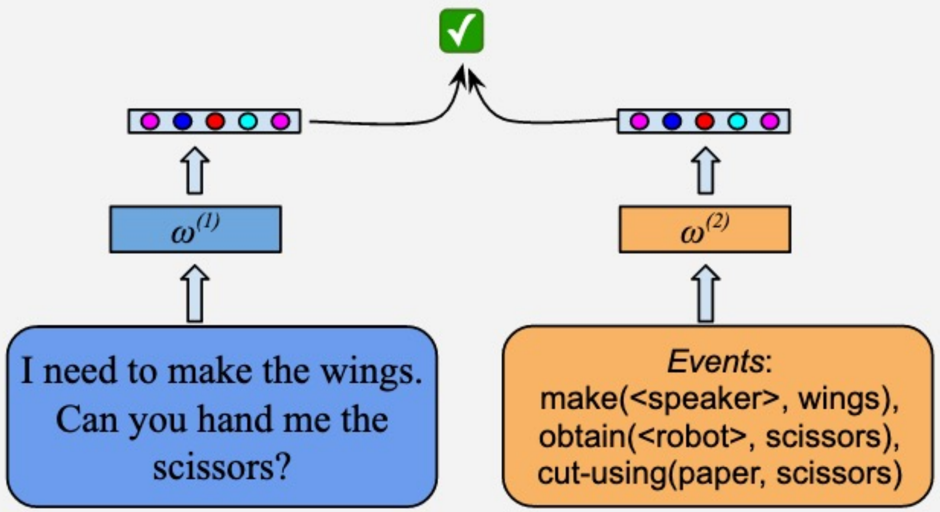


Nguyen et al., Practical Cross-modal Manifold Alignment for Grounded Language. In prep. arxiv.org/abs/2009.05147

NSF #2024878, UMBC, Cynthia Matuszek — cmat@umbc.edu

Implicit task/event prototypes from language

- Implicit expectations: unstated but known constraints
  - Understand critical but unspecified events
  - Predict items/tools likely to be needed
- Capture "implicit prototypes" in grounded language representations
- Use knowledge from meaning representations to improve semantic inference



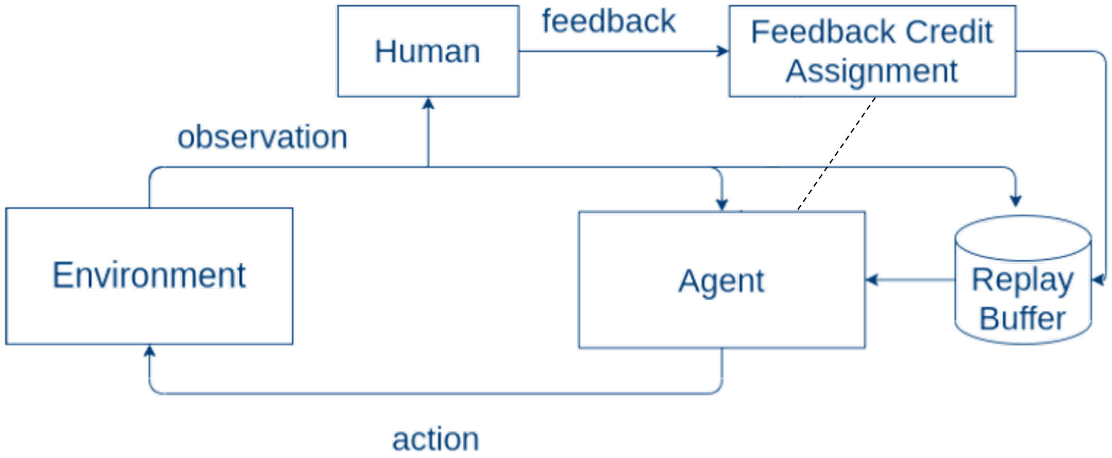
Task-Agnostic Alignment: Learn encodings for the raw text and structured representations that are "close" to one another in embedding space.

After alignment, raw text encoder can improve semantic predictions

Umair & Ferraro, Transferring Semantic Knowledge via Manifold Alignment. Under submission, ACL 2021.

Transferring across robot platforms

- Don’t want every robot to learn every task from scratch
  - Agent-to-agent transfer of learned tasks and world models needed
- Tasks, settings, and robots are never perfectly identical
  - Needs abstractions of learned world / task models
  - Task-based domain adaptation
- Treat as a policy learning effort



Interactive reward shaping for reinforcement learning:

1. Learn task-specific concepts from latent representations of natural language
2. Implicit language learning for domain adaptation
3. Use lifted abstract (hierarchical) MDPs for agent-to-agent transfer

Richards, Winder, & Matuszek, Interactive Reinforcement Learning for Grounded Language Acquisition of Tasks. In prep.

Impacts

- Scaling of language-based collaborative robotics
  - Richer language understanding
  - Sharing learned world and task models
- Customizability of robots in many settings
- Improved multimodal, contextual learning

- Extend robot learning to less represented groups
  - Learning directly from end users improves representation
  - E.g., items from non-Western homes, idiosyncratic task performance, recognition of accented speech