LANGUAGE-CONDITIONAL IMITATION LEARNING

Julian Skirzyński[1], Bobak Baghi[2], David Meger[2]

[1]University of California, San Diego, [2]McGill University

NAACL-HLT 2021: Visually Grounded Interaction and Language (ViGIL) Workshop, Virtual

Abstract

- **Overview**: Introducing Language-Conditional Imitation Learning algorithm (L-CIL) that uses natural language to guide behavior of artificial agents
- **Method**: Language reproduction with behavioral cloning
- **Implementation**: Neural network
- **Results**: Success with multiple behaviors and unseen behaviors; Issues with behavioral cloning
- **Implications**: Simple and promising direction for robotics

Background

- Imitation learning: Mimicking transitions in $\mathcal{D} = \{(o_i, a_i)\}_{i=0}^{N}$
- Behavioral cloning: Solving $\operatorname{minimize}_\theta \sum_t \mathcal{L}(\pi_\theta(o_t;\theta), a_t)$
- Conditional Imitation Learning: Latent information in command c_t, solve $\operatorname{minimize}_\theta \sum_t \mathcal{L}(\pi_\theta(o_t, c_t;\theta), a_t)$ [1]

Method

- Input: Trajectories and sentence descriptions of multiple behaviors in a dataset $\mathcal{D} = \{(o_t, s_t, a_t)\}_{t=1}^{T}$
- Transform sentences s_t into word vectors [2] $v_\psi(s_t)$
- Let $\ell_a(x_1, x_2), \ell_s(x_1, x_2)$ be loss functions that compare actions and sentences representations, and let $\chi_i(\mathbf{x})$ denote a projection on i-th dimension. Let $F(\cdot,\cdot;\theta)$ approximate $(o_t, v_\phi(s_t)) \underset{\theta}{\mapsto} (a_t, v_\phi(s_t))$
- Optimize

$$\underset{\theta}{\operatorname{minimize}} \sum_t \ell_a\left(\chi_1\left(F(o_t, v_\phi(s_t);\theta)\right), a_t\right) + \sum_t \ell_s\left(\chi_2(F(o_t, v_\phi(s_t);\theta)), v_\phi(s_t)\right) \tag{1}$$

References

[1] Felipe Codevilla et al. "End-to-end driving via conditional imitation learning". In: *2018 IEEE International Conference on Robotics and Automation*. IEEE. 2018, pp. 1–9.

[2] Tomas Mikolov et al. "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*. 2013, pp. 3111–3119.

Implementation

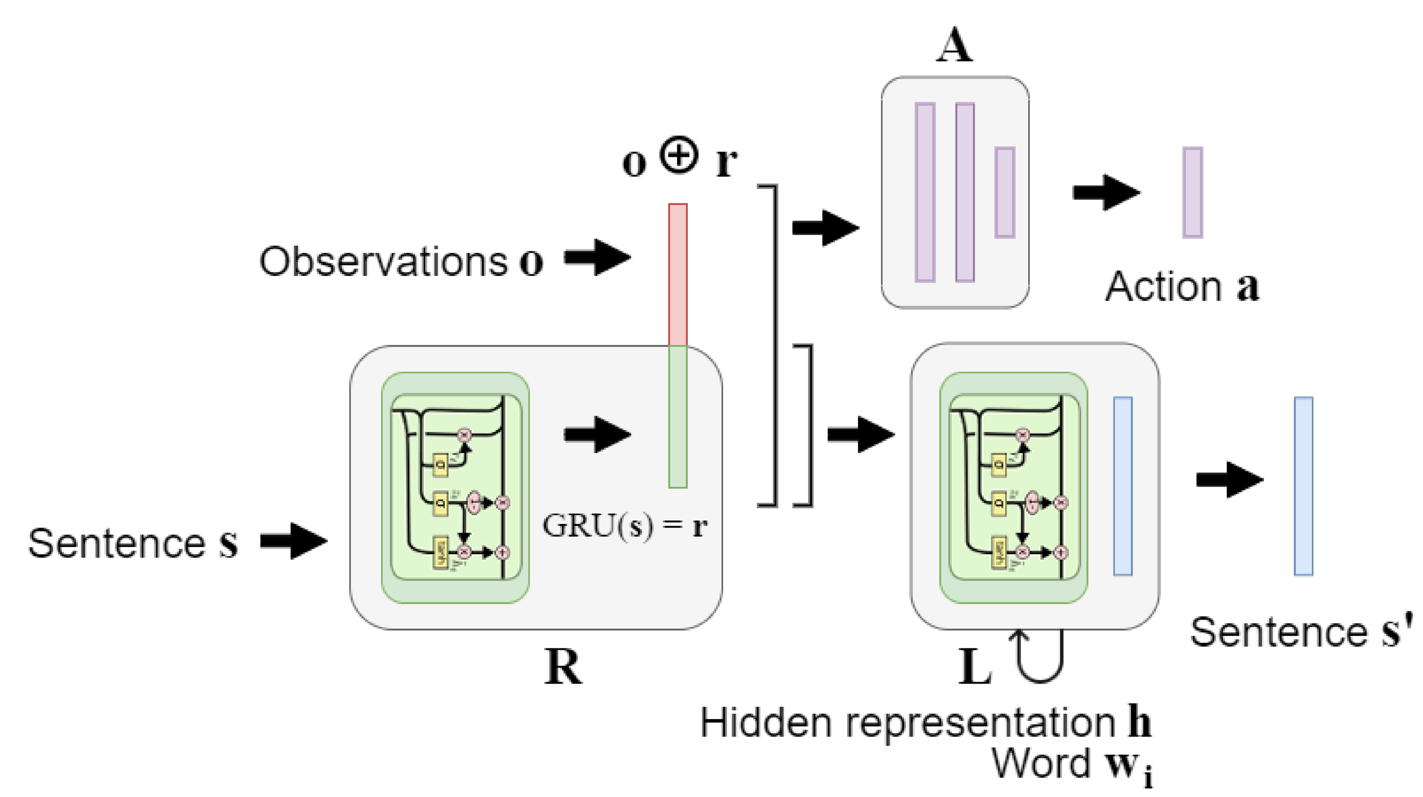

Fig. 1: Network architecture for L-CIL. Image of the GRU taken from https://colah.github.io

- Representation module R maps language to context (encoder)
- Language module L decodes context to language (decoder)
- Action module A maps observations conditioned on context to actions (feed-forward layers)

Experiments

- Driving imitation tasks developed in a self-driving simulator
- Three experiments: imitating multiple behaviors (**MC**), imitating multiple long behaviors (**CC**), imitating unseen behavior knowing the language (**CA**)

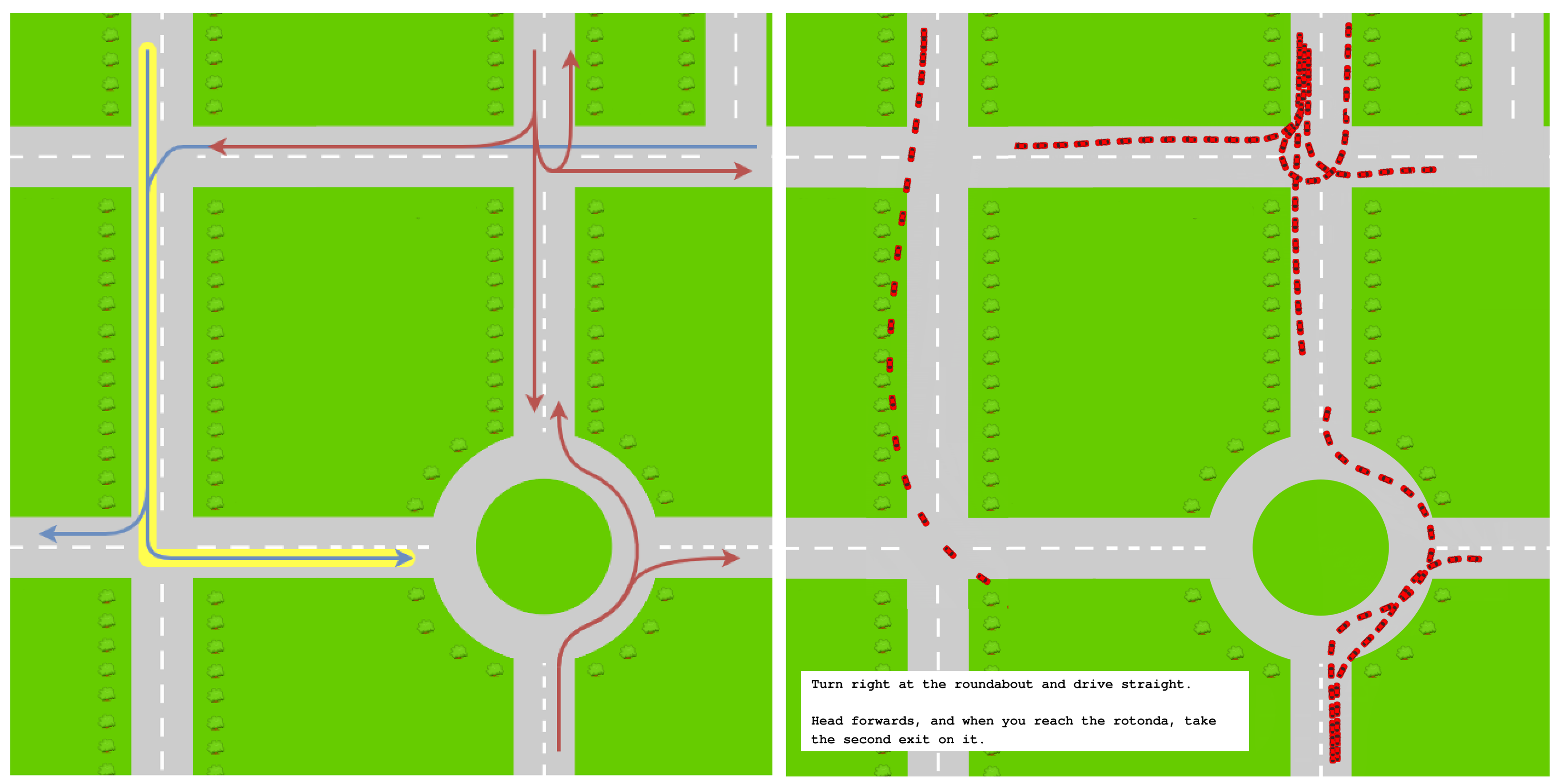

Fig. 2: Left: Map for the experiments with sample trajectories. Right: Sample rollouts outputed by L-CIL.

Baselines and settings

- Behavioral cloning (BC),
- Conditional imitation learning (CIL),
- Language-conditional imitation learning (L-CIL)
- Encoder language-conditional imitation learning (EL-CIL): $\operatorname{minimize}_\theta \sum_t \ell_a\left(\chi_1\left(F(o_t, v_\phi(s_t);\theta)\right), a_t\right)$
- From 2 to 6 behaviors with 100 trajectories each and over 600 000 sentences of length from 11 to 31 in total.

Results

	Experiment		
Algorithm	MC	CC	CA
BC	0.062*	0.014*	**0.028**
CIL	0.021	**0.008**	1.064*
EL-CIL	**0.017**	0.016*	0.101*
L-CIL	0.029*	0.015*	0.033

* difference to the lowest, bolded value is significant with p<0.05

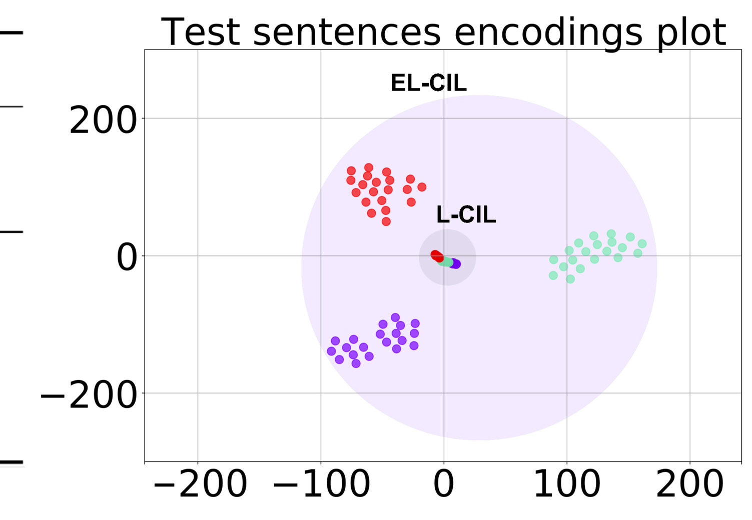

Fig. 3: Left: Mean error for different experiments and algorithms. Right: Test sentence embeddings for EL-CIL and L-CIL

- L-CIL generalizes: improvement over EL-CIL and CIL in the **CA** experiment, similar performance across all experiments
- L-CIL fell short to BC in the **CA** experiment
- CIL is best in discrimination experiments, but not much better than L-CIL
- L-CIL generalizes because the sentence embeddings preserve the similarities between the sentences

Discussion

- L-CIL succeeds due to its architectural setup
- L-CIL is a promising direction for Human-Computer Interaction or robotics research
- Further studies should improve the **CA** experiment