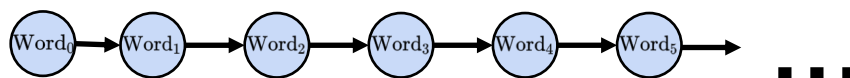


Addressing Optimism Bias in Sequence-Modeling for Reinforcement Learning

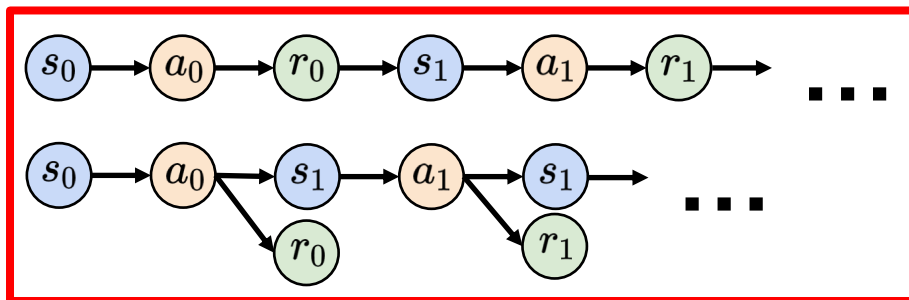
Adam Villaflor, Zhe Huang, Swapnil Pande, John
Dolan, Jeff Schneider

Sequence-Models for RL

- Treat Offline RL as a single sequence modeling problem
 - Analogous to NLP



Word Sequence

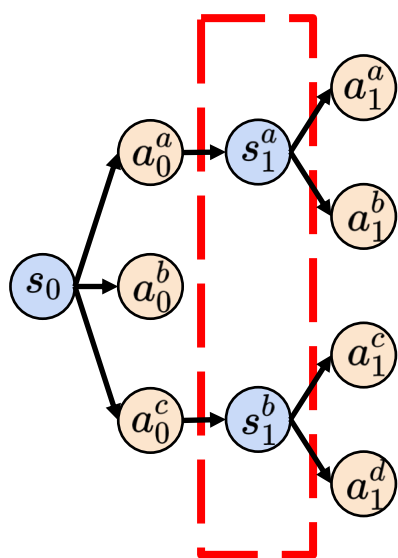


RL Trajectory Sequence

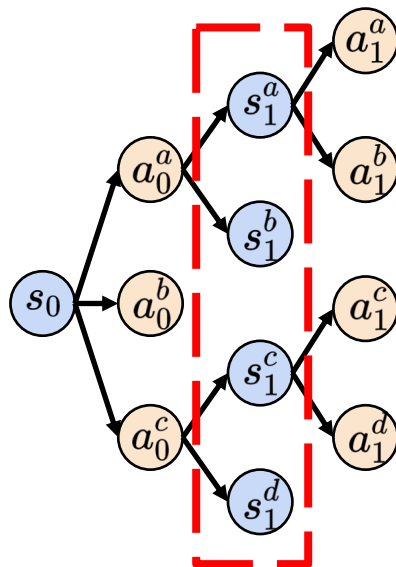
Janner, Michael, Qiyang Li, and Sergey Levine. "Offline reinforcement learning as one big sequence modeling problem." *Advances in neural information processing systems* 34 (2021): 1273-1286.

Chen, Lili, et al. "Decision transformer: Reinforcement learning via sequence modeling." *Advances in neural information processing systems* 34 (2021): 15084-15097.

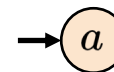
RL Search Trees Issue



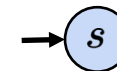
Deterministic Environment



Stochastic/Adversarial Environment



Agent Controlled



Environment Controlled

2 Possibilities Dictated by Environment

**Speed
Through**



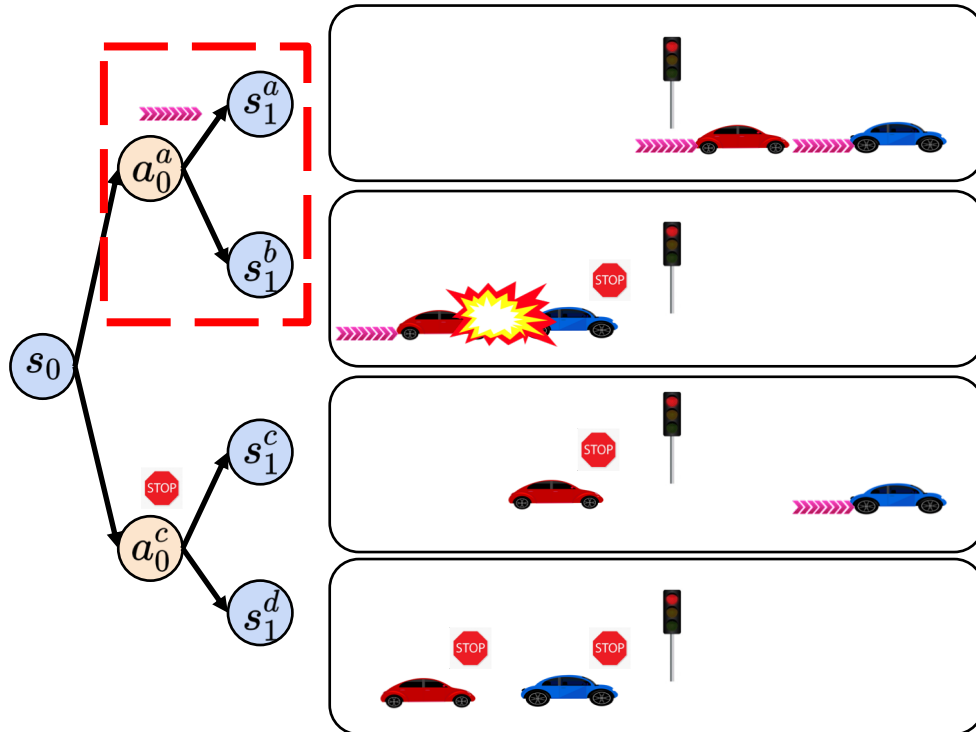
**Slow to
a Stop**



Ego Agent

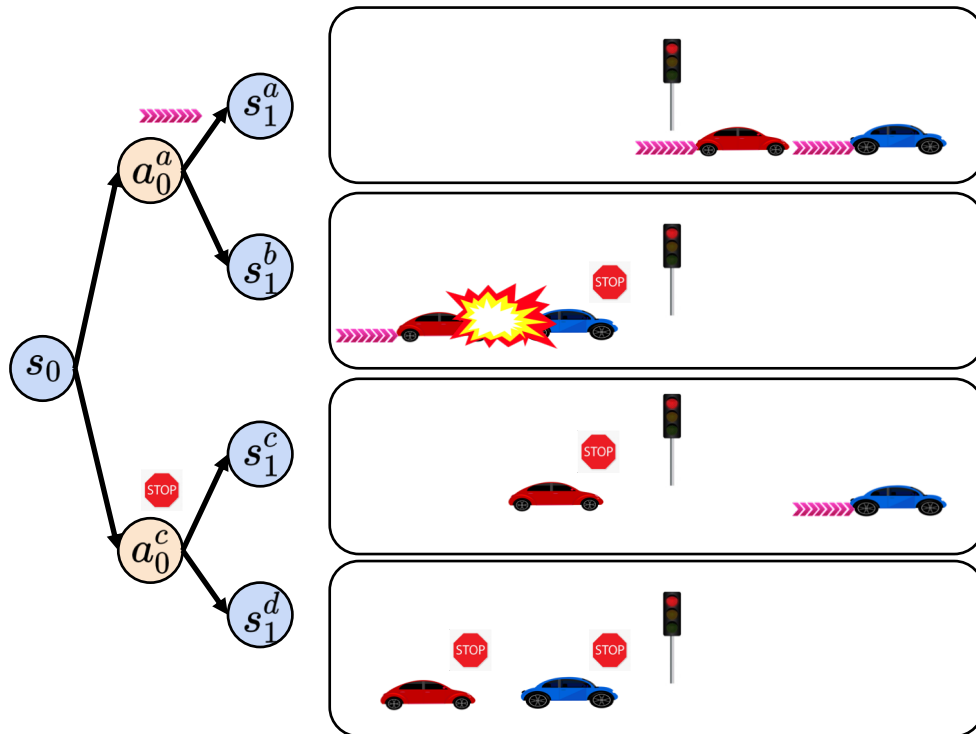


RL Search Trees Issue



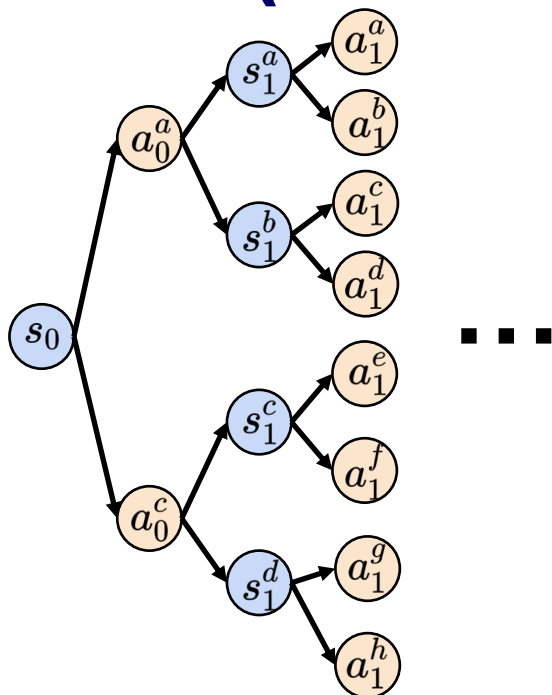
- How do you handle state branching?
 - Standard RL: Expectation
 - Adversarial Game/Robust RL: Minimum
- What does naive beam search do?
 - Analogous to maximum

Return-Conditioning Issue

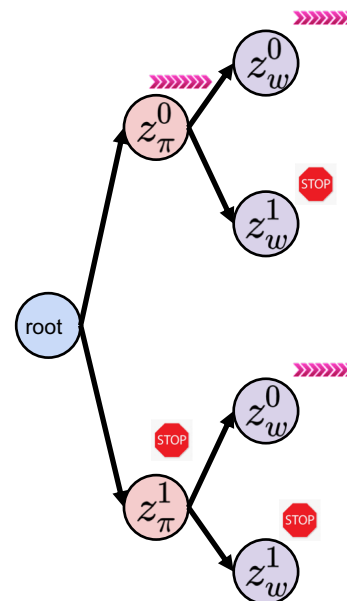


- What is the correct target return?
 - High return → Aggressive
 - Low return → Cautious

SeParated Latent Trajectory (SPLT) Transformer

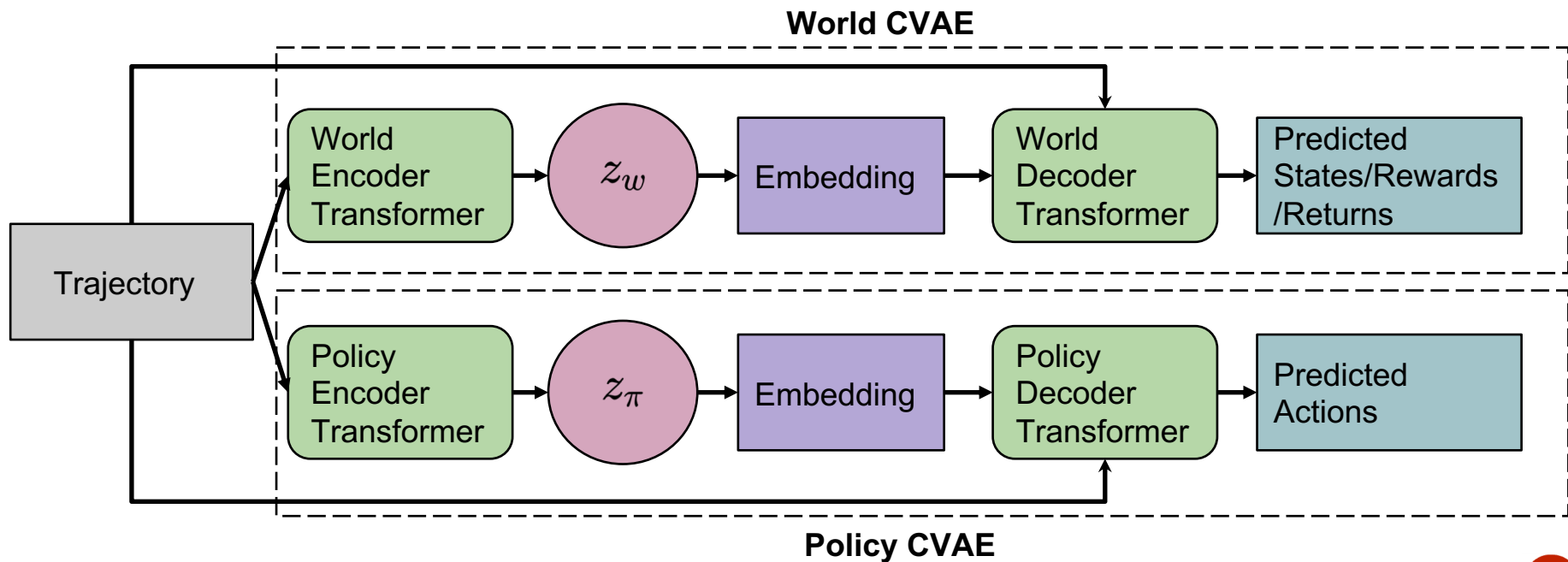


Full Step-Wise Tree Search

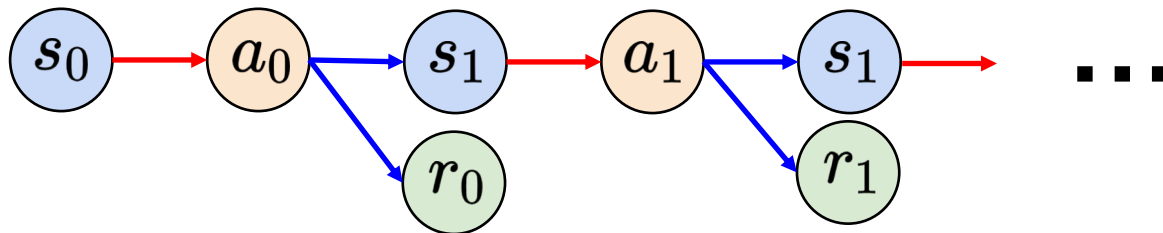


Our Latent Variable Search

SPLT Transformer

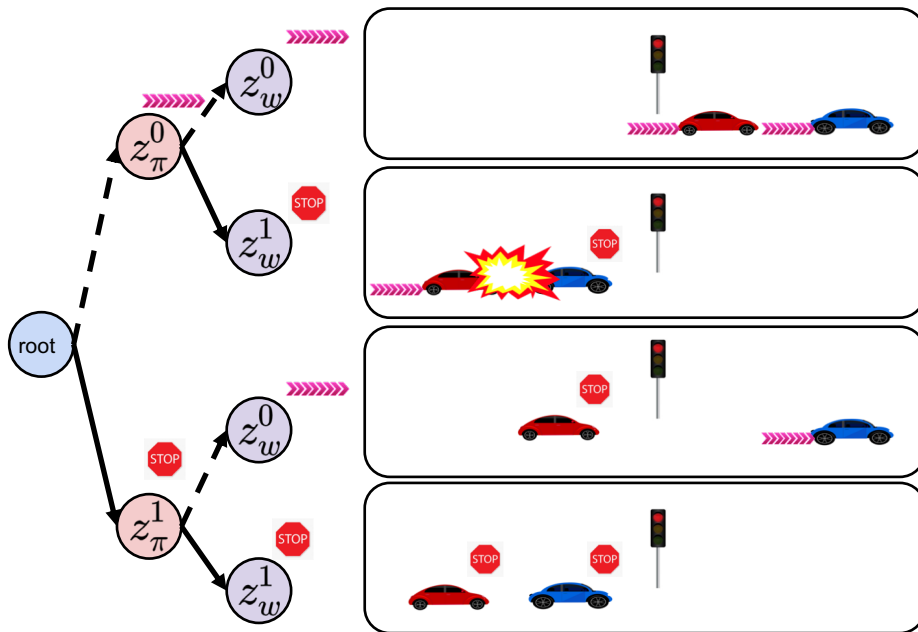


SPLT Transformer Trajectory Generation



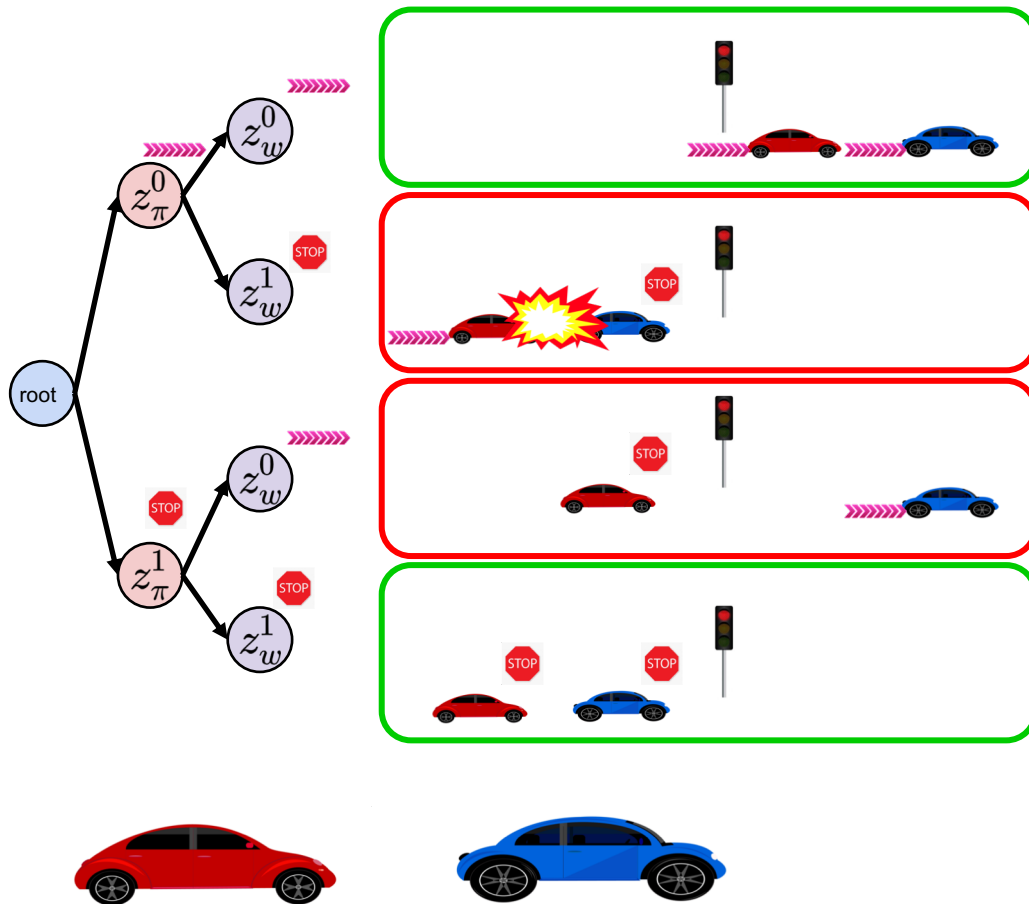
RL Trajectory Sequence

SPLT Transformer Search



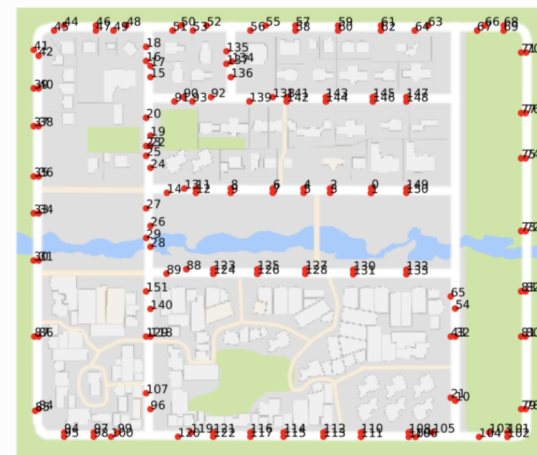
- Evaluate using predicted return for each predicted trajectory
- Take min over world latent variables
- Take max over policy latent variables

Our Latent Variable Search



Carla Results

Metric	SPLT (Ours)	BC	TT	DT(m)
Success (%)	96.3 ± 2.1	95.3 ± 0.4	89.0 ± 2.8	92.0 ± 7.1
Speed (m/s)	2.71 ± 0.04	2.51 ± 0.0	2.67 ± 0.1	2.72 ± 0.03



Codevilla, Felipe, et al. "Exploring the limitations of behavior cloning for autonomous driving." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Dosovitskiy, Alexey, et al. "CARLA: An open urban driving simulator." Conference on robot learning. PMLR, 2017.