Lightweight and Faithful Visual Condition Checking in Behavior Trees via Expert-Regularized Reinforcement Learning

Hyosik Moon[1], Eldan Cohen[1] | [1] University of Toronto, Toronto, Canada

Association for the Advancement of Artificial Intelligence

Motivation

- **Behavior Trees** (BTs) are interpretable, but traditional BTs rely on hand-crafted symbolic conditions, struggling with visual and textual inputs.
- **Vision-Language Models** (VLMs) enable to process high-dimensional perceptual inputs, but they are too slow and expensive for real-time use.
- To build lightweight alternatives, **Imitation Learning** (IL) can be used but requires large amounts of expert labels, while **Reinforcement Learning** (RL) avoids expert supervision but often causes semantic drift and poor credit assignment in BTs.

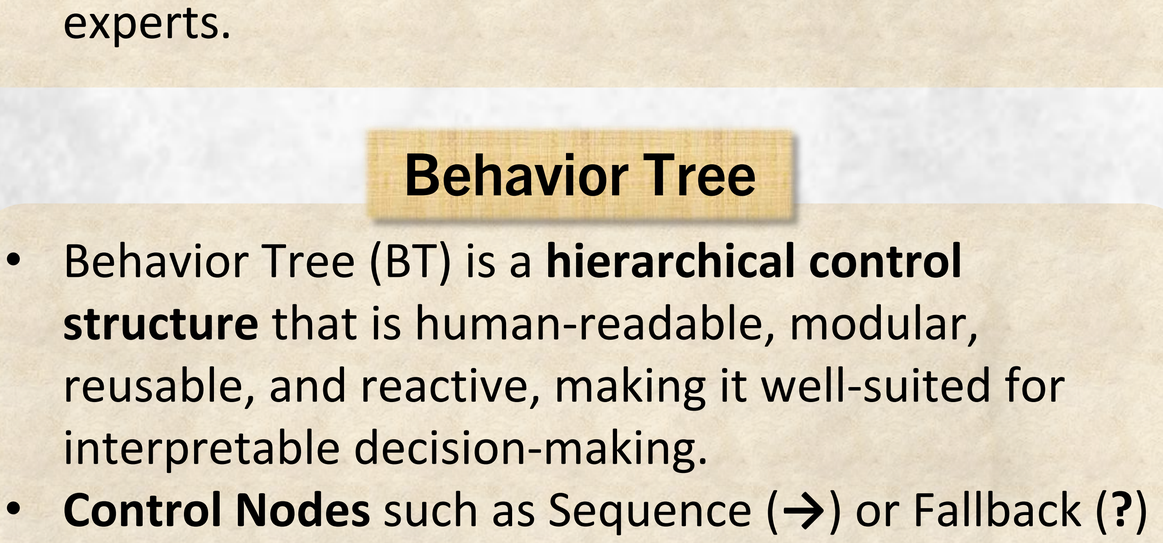
We need a lightweight and faithful way to check conditions from perceptual inputs in BTs.

Contribution

- A unified framework **integrating IL and RL** via expert regularization and **sampling-based policy gradients** for BT-driven polices from perceptual inputs.
- **Factorized formulation** that aggregates a sequence of condition-node decisions into a single decision unit to improve credit assignment.
- Achieves high success rates over IL or RL alone, strong agreement with expert decisions, and orders-of-magnitude faster inference than VLM experts.

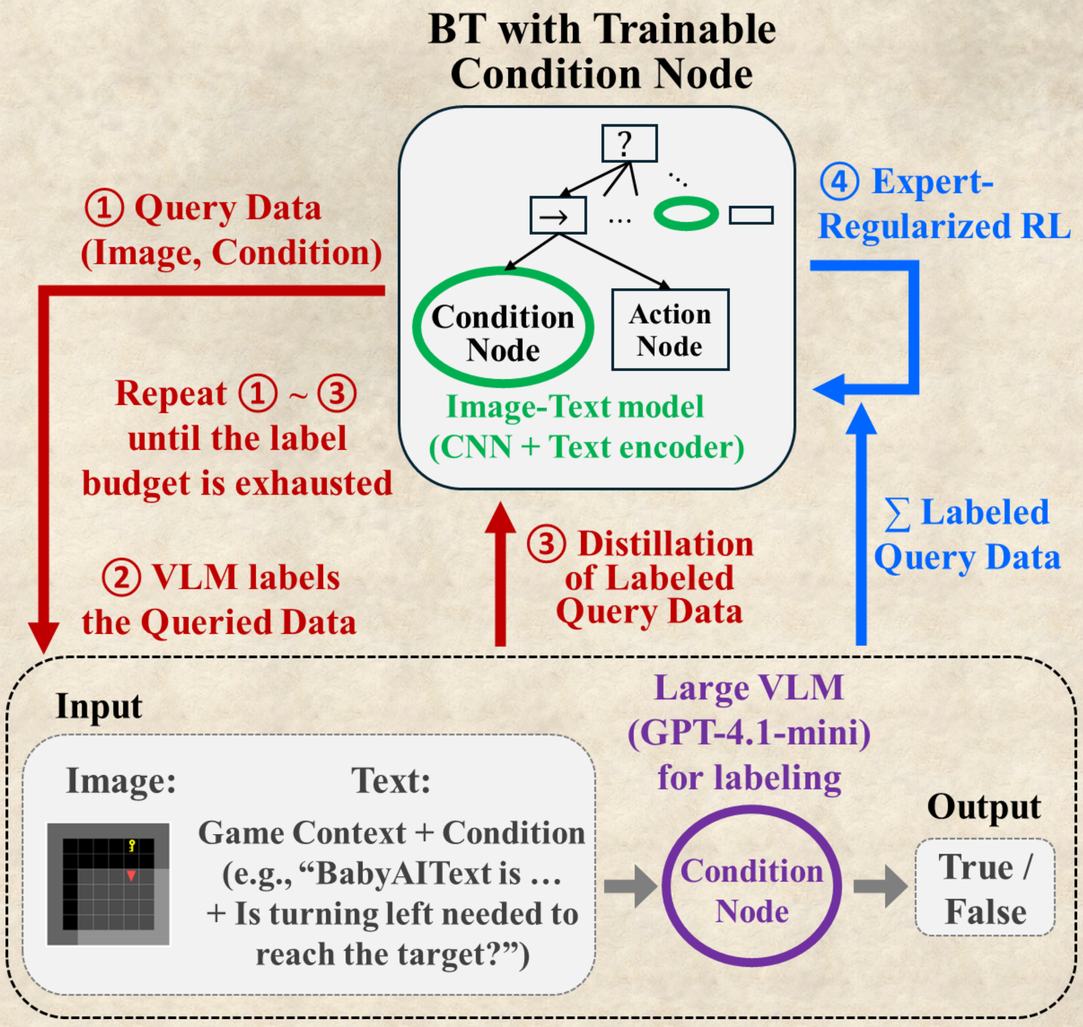
Behavior Tree

- Behavior Tree (BT) is a **hierarchical control structure** that is human-readable, modular, reusable, and reactive, making it well-suited for interpretable decision-making.
- **Control Nodes** such as Sequence (→) or Fallback (**?**) manage the logic flow, while **Leaf Nodes** perform Actions (rectangles) or check Conditions (ellipses).
- A "**tick**" signal pulses from the root to leaf, allowing the tree to be reactive by constantly updating node statuses as Running, Success, or Failure.



e.g., Behavior tree for BabyAIText (Pickup)

Methodology



- **Expert Labeling (①, ②):** A VLM (GPT-4.1-mini) evaluates natural-language condition queries on image inputs.
- **Warm Start (③):** The lightweight condition node is initialized using Behavior Cloning or DAgger.
- **Expert-Regularized RL (④):** PPO is used to optimize task reward while maintaining semantic alignment via cross-entropy regularization with the expert labels used during initialization.

Experiment Results

Metrics	Model		Success Rate	Accuracy
Avg (%) ↑	Expert (GPT-4.1-mini)		84.89	100.00
	IL	BC	59.75	90.35
		DA	50.32	89.53
	RL	RL_b	19.14	63.10
		RL_f (ours)	47.29	67.51
	IL + RL (ours)	$init_{BC}$ + RL_f w/ $ER_{1.0}$	77.96	**91.89**
		$init_{BC}$ + RL_f w/ $ER_{0.1}$	**79.89**	89.50
	Ablation	RL_f w/ $ER_{0.1}$	52.75	83.44
		$init_{BC}$ + RL_f	77.21	83.34
		$init_{DA}$ + RL_f w/ $ER_{0.1}$	74.04	90.22

- Evaluated on **7 tasks** from GymCards, FrozenLake, and BabyAIText over 4 random seeds.
- **Performance**: IL+RL achieves ~80% success rate, outperforming IL (~60%) and RL (~47%) baselines.
- **Accuracy** : Maintains ~90% agreement with expert
- **Efficiency**: Inference time, ~0.09s/episode vs ~63s/episode; Model size, 6.6M vs ~7B parameters , for IL+RL models and VLM respectively.
- **Cost**: ~208k expert queries (~$97 total).