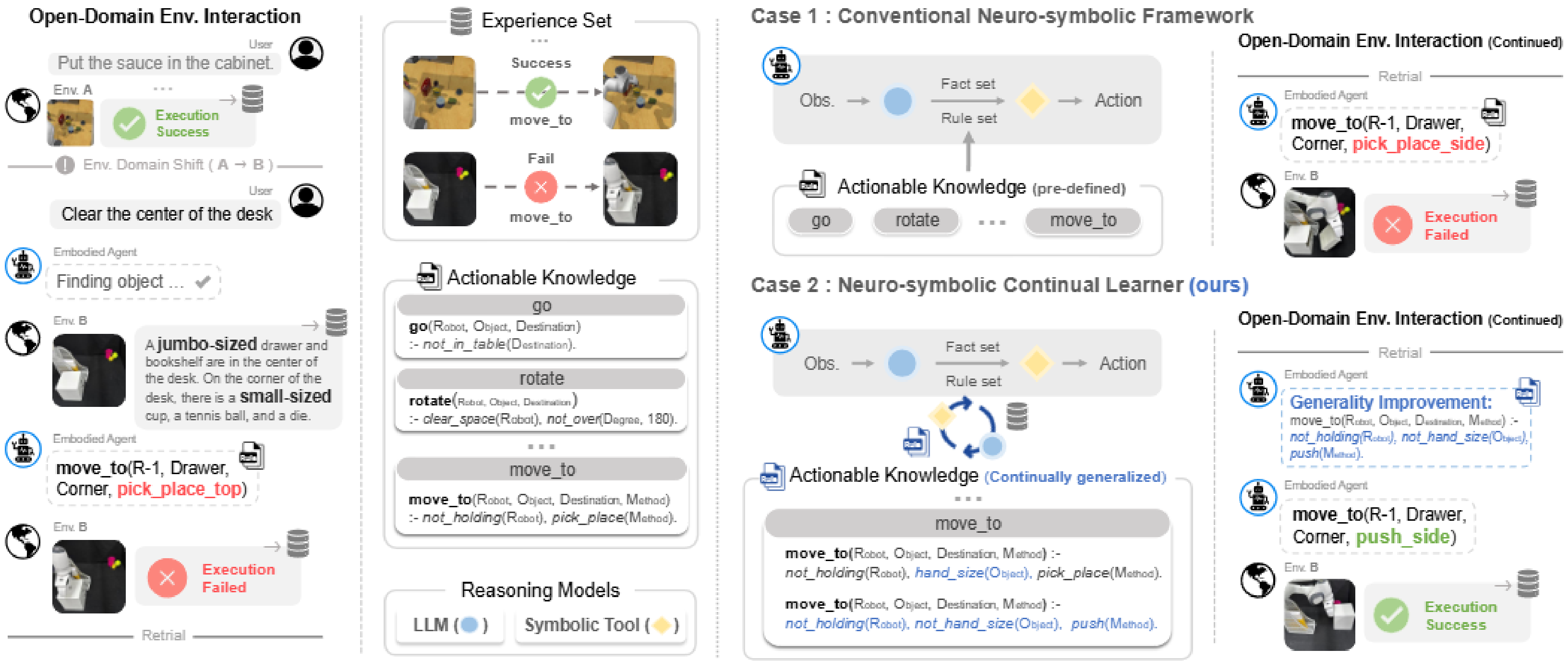
NeSyC: A Neuro-Symbolic Continual Leaner For Complex Embodied Tasks In Open Domains

Wonje Choi*, Jinwoo Park*, Sanghyun Ahn, Daehee Lee, Honguk Woo

Department of Computer Science and Engineering, Sungkyunkwan University

{wjchoi1995, pjw971022, shyuni5, dulgi7245, hwoo}@skku.edu

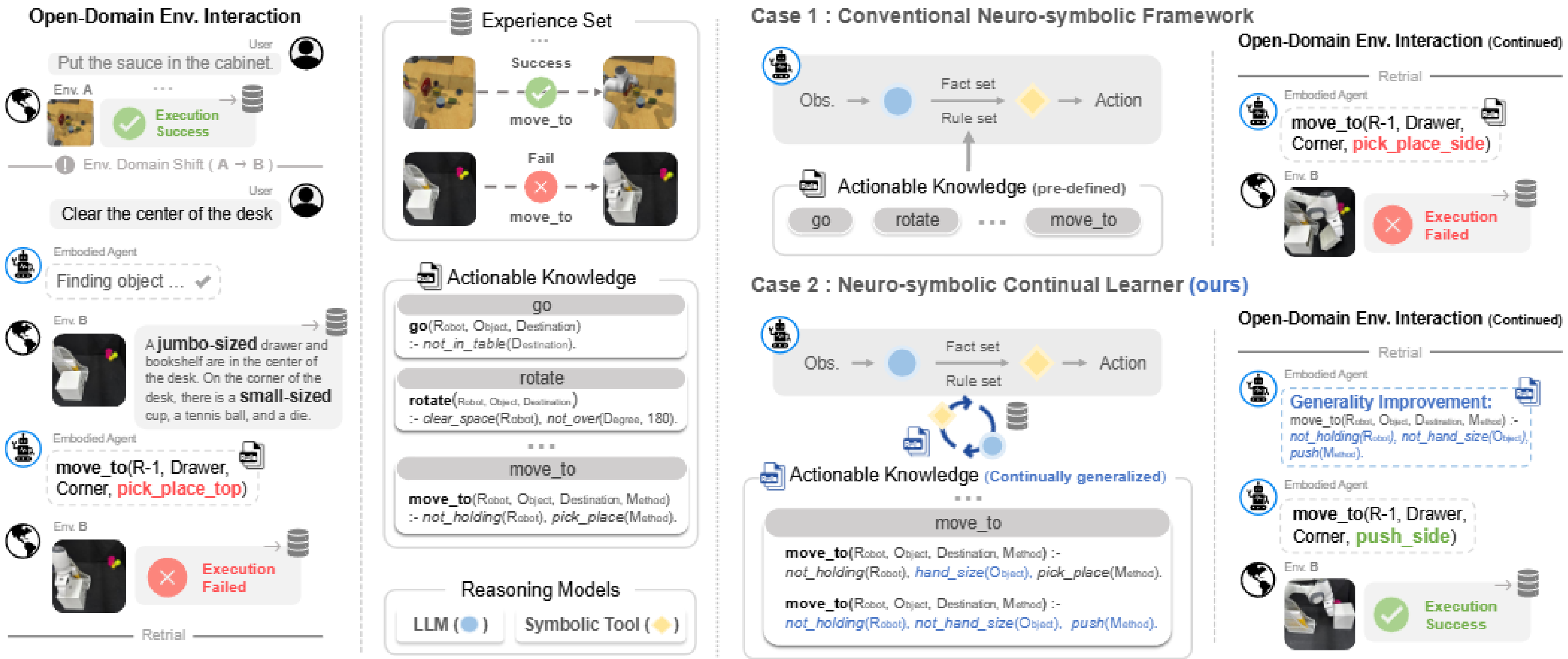
Introduction



1. Neuro-symbolic systems for embodied tasks

- integrating Large Language Models (LLMs) with symbolic tools for planning (e.g., ASP solver, PDDL solver).
- existing neuro-symbolic approaches rely on predefined, fixed rules tailored to specific tasks or environments.

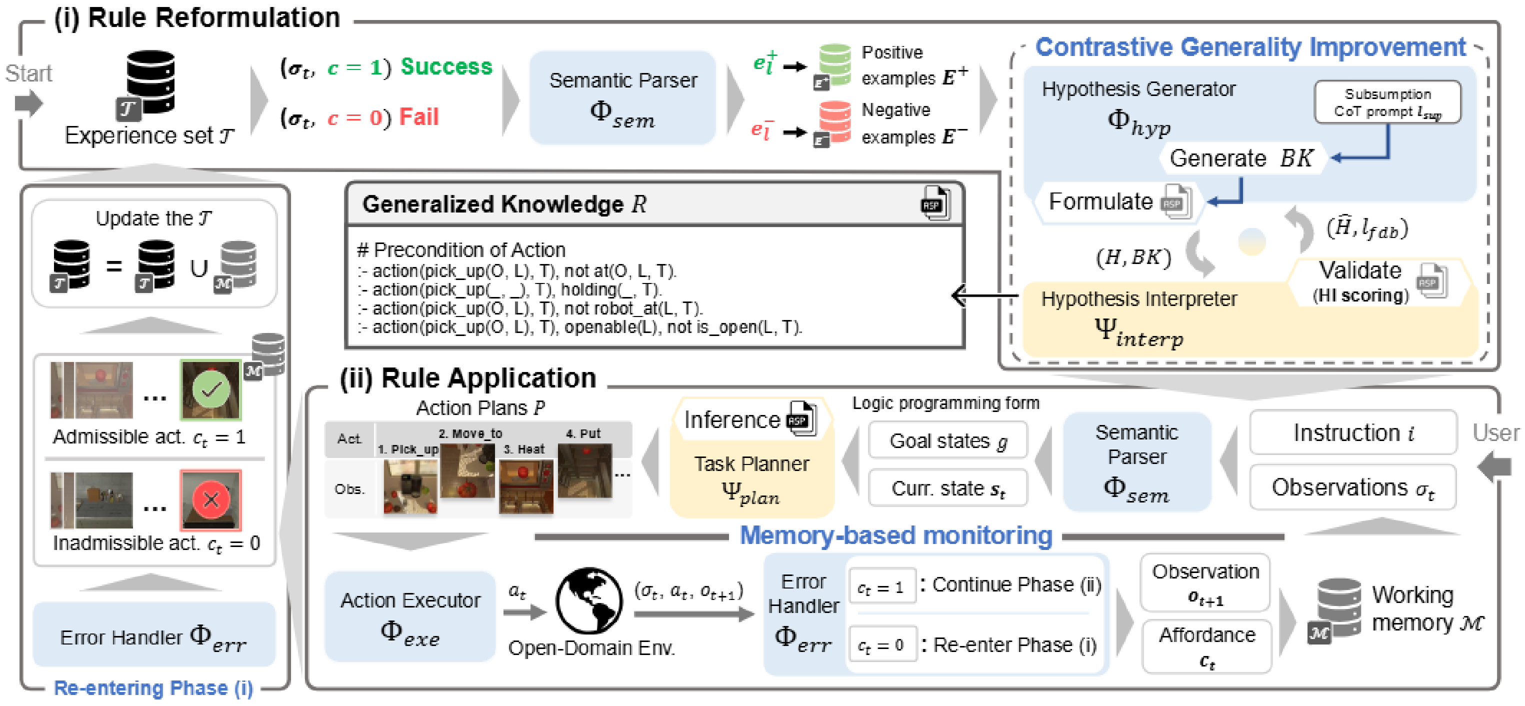
Introduction



2. Generalization of actionable knowledge in open-domain environments

- Our goal is to generalize actionable knowledge using limited agent’s experiences for continual tasks.
- Drawing inspiration from the hypothetico-deductive model, continually forming hypotheses, rigorously testing them against available observations, and iteratively revise them.

Neuro-Symbolic Continual Leaner



1. NeSyC framework is designed:

- To leverage LLMs for rich language-based reasoning and hypothesis generation
- To employ symbolic tools for logical validation and error detection

Evaluation: Setting

- Environment
  - For simulation experiment, we use ALFWorld, VirtualHome, Minecraft, RLBench
  - For real-world scenario, we design table-top rearrangement tasks
- Baselines
  - LLM-based planning methods: LLM-planner
  - Multi-agent frameworks: ReAct and Reflexion, AutoGen
  - Neuro-symbolic approaches: ProgPrompt and CLMASP
- Metrics
  - Task Success Rate (**SR**) (%) is the percentage of tasks fully completed
  - Goal-conditioned Success Rate (**GC**) (%) is the percentage of sub-goals that are completed.

Evaluation: Performance

ALFWorld	Static			Low Dynamic			High Dynamic		
Methods	SR	GC	Step	SR	GC	Step	SR	GC	Step
LLM-planner	10.6±2.8	17.7±3.4	20.9±3.7	9.8±2.7	22.2±3.8	26.8±4.0	7.3±2.4	17.1±3.4	21.1±3.7
ReAct	35.8±3.5	48.2±4.1	51.7±4.3	34.1±4.3	45.2±4.5	50.6±4.5	18.7±3.5	28.1±4.1	33.5±4.3
Reflexion	39.0±3.7	63.5±4.4	67.0±4.5	37.4±4.4	64.8±4.3	70.6±4.1	21.1±4.4	41.5±4.3	43.6±4.2
AutoGen	58.5±4.4	77.8±3.7	81.1±3.5	51.2±4.5	69.3±3.9	75.6±4.2	30.9±4.2	50.6±4.5	58.8±4.4
CLMASP	**88.5±2.9**	**89.1±2.8**	**89.1±2.8**	58.8±4.4	63.3±4.4	71.8±4.1	23.8±3.9	36.5±4.4	45.8±4.5
NeSyC	82.9±3.4	83.5±3.4	83.6±3.3	**78.9±3.7**	**79.4±3.7**	**80.0±3.6**	**70.7±4.1**	**75.5±3.9**	**76.4±3.8**

VirtualHome	Static			Low Dynamic			High Dynamic		
Methods	SR	GC	Step	SR	GC	Step	SR	GC	Step
LLM-planner	21.5±0.5	33.2±0.4	33.0±9.7	20.5±0.6	32.7±0.8	29.8±2.7	14.8±3.4	27.7±2.9	21.5±2.2
ReAct	40.0±5.0	51.9±4.5	44.8±0.8	34.6±3.6	46.8±3.4	35.9±3.7	17.2±2.1	32.4±1.2	18.8±2.0
Reflexion	36.2±1.7	47.5±2.0	16.4±1.1	35.4±1.9	46.6±1.0	36.5±1.8	15.5±0.9	29.7±1.6	16.4±1.1
AutoGen	44.3±2.2	54.8±2.7	45.8±2.2	43.2±1.2	54.4±1.4	44.9±1.2	18.9±0.9	33.0±1.7	20.6±1.0
CLMASP	76.4±0.8	**89.1±0.8**	84.1±0.0	28.9±0.4	42.5±0.2	28.9±0.4	0.0±0.0	13.1±0.0	0.0±0.0
NeSyC	**82.3±0.4**	87.4±0.6	**84.2±0.6**	**79.6±1.9**	**85.8±1.1**	**80.8±1.9**	**77.5±1.3**	**84.1±0.6**	**79.0±1.2**

Minecraft	Static			Low Dynamic			High Dynamic		
Methods	SR	GC	Step	SR	GC	Step	SR	GC	Step
LLM-planner	31.1±1.5	41.2±1.4	42.5±2.4	28.9±4.2	31.9±1.6	37.0±1.6	23.3±2.7	25.8±1.3	28.9±0.4
ReAct	34.4±1.6	38.3±1.4	44.4±2.7	27.8±1.6	31.6±1.7	40.5±4.0	21.1±1.6	24.7±2.8	30.6±3.6
Reflexion	41.1±1.6	47.2±1.3	49.0±1.6	30.0±2.7	34.0±2.7	39.1±2.3	21.1±1.6	24.0±2.4	30.9±4.0
AutoGen	51.1±4.2	52.2±3.4	53.9±2.8	33.3±2.7	36.6±2.4	38.2±2.5	25.6±3.1	28.3±3.6	32.7±4.2
CLMASP	**94.4±3.1**	**95.4±1.7**	**95.8±1.3**	52.2±5.7	55.7±6.4	59.1±5.5	48.9±3.1	50.9±3.6	52.8±2.8
NeSyC	92.2±1.6	94.3±0.9	95.3±1.0	**91.1±1.6**	**93.2±1.4**	**94.1±1.4**	**87.8±5.7**	**89.9±5.9**	**90.9±5.9**

RLbench	Static			Low Dynamic			High Dynamic		
Methods	SR	GC	Step	SR	GC	Step	SR	GC	Step
LLM-planner	16.7±5.7	23.3±2.9	35.5±1.9	16.7±2.9	20.8±1.4	27.4±1.0	18.3±2.9	21.7±1.4	27.2±1.9
ReAct	23.3±2.9	25.8±1.4	36.5±1.7	21.7±2.8	23.3±1.4	30.8±1.6	18.3±2.9	20.0±2.5	26.8±2.5
Reflexion	33.3±2.8	41.4±5.1	47.5±3.3	21.7±2.9	24.4±2.1	32.1±2.5	23.3±2.8	23.3±2.8	29.6±2.8
AutoGen	43.3±8.6	54.2±4.6	57.9±3.3	23.3±2.9	28.6±2.7	32.1±2.3	21.7±5.8	23.3±2.9	28.9±1.9
CLMASP	**94.5±4.2**	**95.8±2.8**	**96.0±2.7**	0.0±0.0	6.0±0.8	25.0±2.0	0.0±0.0	3.7±0.7	12.3±0.9
NeSyC	85.5±2.7	88.5±0.9	91.9±0.7	**81.5±4.5**	**84.8±4.5**	**88.7±3.6**	**79.0±6.6**	**81.8±7.0**	**86.2±6.2**

Table1. Performance evaluation on simulation environments

- Experiment results

- NeSyC achieves up to a 50% higher success rate compared to multi-agent framework baseline, AutoGen

Evaluation: Performance

Real-world	*Static*		*Dynamic*		**Real-world**	*Static*		*Dynamic*		**Real-world**	*Static*		*Dynamic*	
Binary	SR	GC	SR	GC	*Cause*	SR	GC	SR	GC	*Guidance*	SR	GC	SR	GC
LLM	22.2	59.2	11.1	42.4	LLM	66.6	87.0	44.4	79.7	LLM	55.6	81.4	33.3	72.1
					Human	77.8	92.6	55.6	85.2	Human	88.9	98.1	66.7	94.3

Table2. Performance evaluation on real-world environments

- Experiment results

- NeSyC achieves up to a 50% higher success rate compared to multi-agent framework baseline, AutoGen

Conclusion

- A Novel Framework: NeSyC

- Enabling effective embodied task planning in open domains by continually generalizing actionable knowledge from experiences

- Future Direction

- Exploring neuro-symbolic knowledge distillation for resource-efficient embodied control with smaller LLMs.