Leveraging Temporal Reasoning for Policy Selection in Learning from Demonstration

Estuardo Carpio, Madison Clark-Turner, Paul Gesel, Momotaz Begum

{erp48, mbc2004, pac48, mbegum}@cs.unh.edu

University of New Hampshire

Cognitive Assistive Robotics Lab

Introduction

High-level human activities often have rich temporal structures that determine the order in which actions are executed. We propose the **Temporal Context Graph (TCG)** to capture these temporal structures. We present Learning from Demonstration (LfD) as the application domain in which the use of TCGs can improve policy selection and address the problem of perceptual aliasing.

Background

- **Motivation:** address shortcomings of previous approaches when dealing with perceptual aliasing and cyclical actions.
- **Allen's Interval Algebra:** Describes the interval temporal relations (ITR) that can exist between two actions that take place over an interval of time.

Allen's interval temporal relations

Relation	Notation	Graphical Representation	Inverse
Before	X{b}Y	X Y	Y{bi}X
Meets	X{m}Y		Y{mi}X
Overlaps	X{o}Y		Y{oi}X
Starts	X{s}Y		Y{si}X
During	X{d}Y		Y{di}X
Finishes	X{f}Y		Y{fi}X
Equals	X{eq}Y		Y{eq}X

- **Temporal Context:** The set of actions that have taken place from the start of the execution of a task to the current point in time and the ITRs between them.

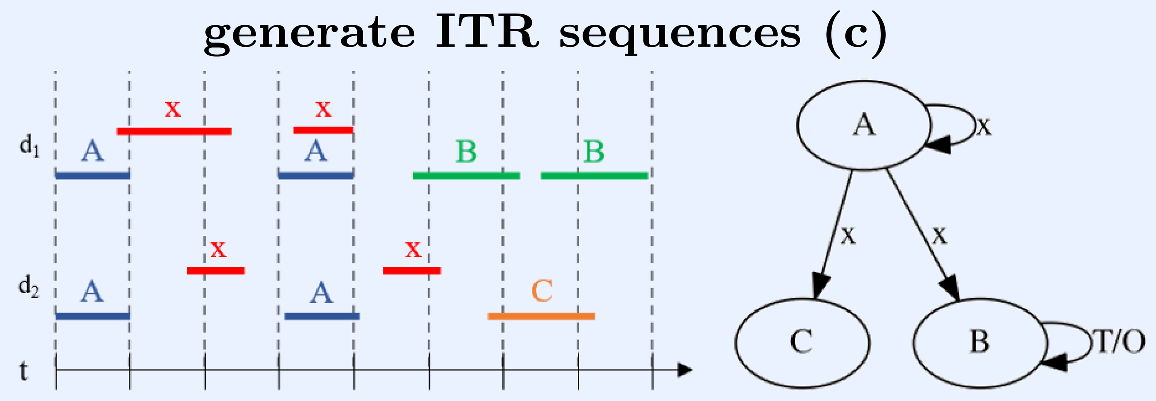
Temporal Context Graph

- TCGs are interval-based temporal reasoning models capable of encoding the temporal structure of task.
- A TCG can be defined as:

$$TCG = \{N, E, P\}$$

  - N is a set of nodes representing the states of the task,
  - E is a set of edges that represents the events that trigger transitions between two states,
  - and P is an n-gram model that encodes the state transition probabilities between two states given the current temporal context.
- A TCG can be learned from a demonstration set:
  - Each demonstration is the set of actions that take place, their respective start and end times, and a flag indicating if they trigger a state transition.
  - Transition and non-transition actions create the edges and nodes of the TCG, respectively.
  - Timing information is used to generate ITR sequences that describe each demonstration.
  - An n-gram model is trained with the set of ITR sequences to learn the temporal contexts of the task.

Toy demonstration set (a) used to learn a TCG (b) and generate ITR sequences (c)



(a) (b)

(A,o,x), (x,b,A), (A,fi,x), (x,b,B), (B,b,B) ⟶ (A,1,x), (x,2,A), (A,3,x), (x,4,B), (B,5,B)

(A,b,x), (x,b,A), (A,b,x), (x,b,C) ⟶ (A,1,x), (x,2,A), (A,7,x), (x,8,C)

(c)

Policy Selection using a TCG

- When used for task execution, a TCG keeps track of its state, which is defined as:

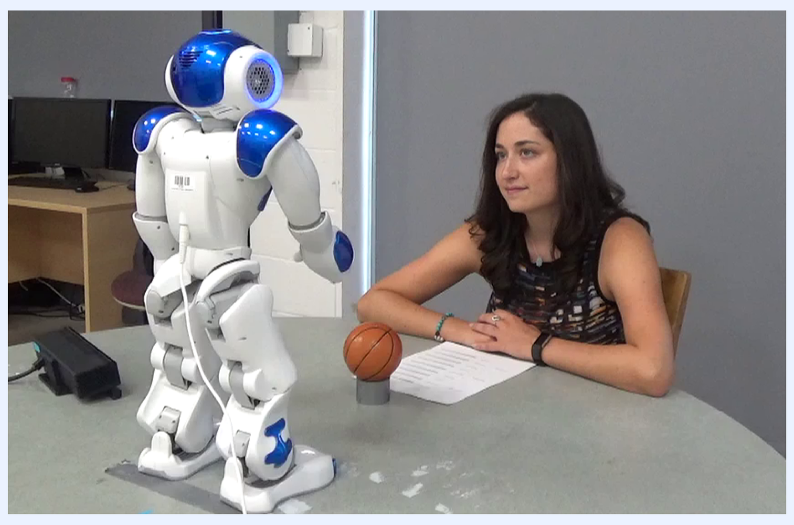
$$S_t = \{n_t, \omega_t, \epsilon, c\}$$

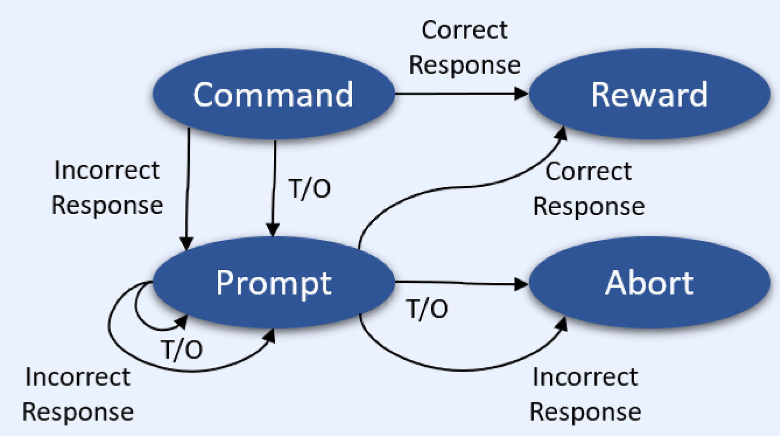
  - n_t represents the current state of the task,
  - ω_t is the waiting period before a timeout is needed,
  - ϵ is the observation of the current state of the environment,
  - and c is an ITR sequence describing the current temporal context of the task.

Results: Object Naming Intervention

- Robot-mediated educational intervention designed to improve the vocabulary of children with autism spectrum disorder (ASD).
- Use case with cyclical actions, as the robot can deliver a *prompt* multiple times.
- Multiple instances of perceptual aliasing per session, as incorrect human responses may be identical before or after a *prompt* is delivered.
- Evaluated on 60 sessions with a success rate of 97%.

Physical setup and TCG learned for the intervention

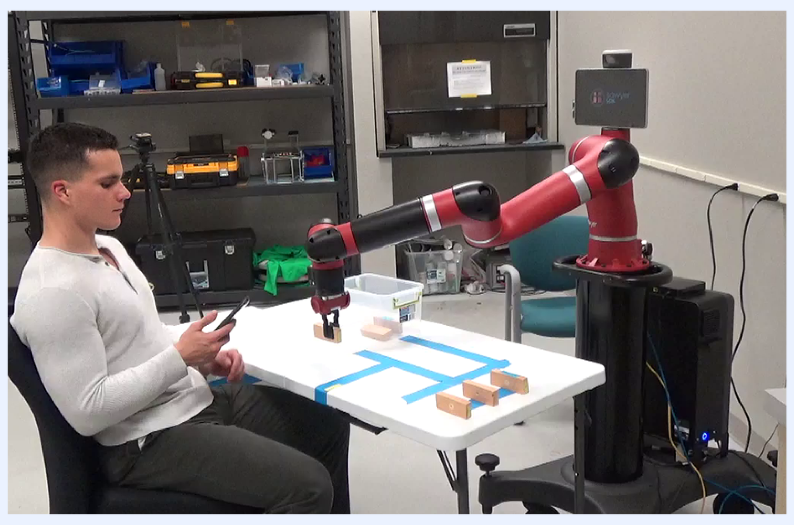


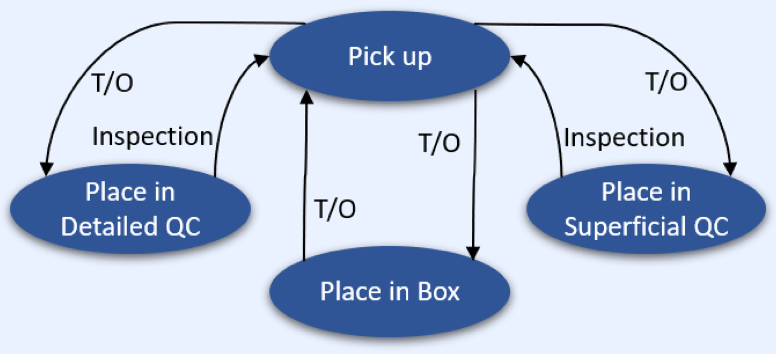


Results: Collaborative Packaging Task

- Collaborative task designed to replicate an industrial environment.
- The *pick up* action is cyclical and must be performed exactly 6 times.
- Policy selection depends on non-sequential ITRs for the *place* actions.
- Perceptual aliasing is present, as the only inputs are the joint states and the QC notifications.
- Policy selection success rate of 100% over 20 evaluation sessions.

Physical setup and TCG learned for the task





Conclusion

- TCGs are interval-based temporal reasoning models capable of learning temporal structures with cyclical actions.
- TCGs can solve instances of perceptual aliasing by leveraging the current temporal context of the task.
- Temporal reasoning can be used in LfD frameworks to limit the action-space of a robotic agent and improve policy selection.

Acknowledgements

Research support

IIS 1664554 NSF

Travel support

POSITION IMAGING