Selecting Subgoals using Deep Learning in Minecraft: A Preliminary Report

Dave Bonnano¹, Mark Roberts², Leslie Smith³, & David W. Aha³

¹Naval Research Laboratory, Code 5557; Washington, DC
²NRC Postdoctoral Fellow; Naval Research Laboratory, Code 5514; Washington, DC
³Naval Research Laboratory, Code 5514; Washington, DC
{david.bonanno,mark.roberts.ctr,leslie.smith,david.aha}@nrl.navy.mil

Presenter: Mark “Mak” Roberts
http://makro.ink

Research Sponsored by NRL
Alex's Goal-Task Network

- Complete Obstacle Course
- Gather Resources
- Survive

- Stay Full
- Defend

Alex
Desirable Research Properties

- Open & uncertain world
- Limitless goals
- Inherent tradeoffs
- Long reward sequences
- Multi-player and multi-agent
- Assistive agents
- Easy to abstract planning and decision making
Alex’s Goal-Task Network

Controller: Serializes actions and disregards unsafe movement
Tradeoffs during execution

Get the diamonds on left!  Avoid the zombies on the left!
How do I choose subgoals during execution?

Goal Reasoning using Goal Lifecycle + Goal Memory

Select subgoals with Deep Learning
Outline

• Goal Reasoning in a Nutshell
• ActorSim implements Goal Reasoning Theory
  – Goal lifecycle & Refinement Strategies
  – Walk through demonstration
• Learning effective decisions for Minecraft
• Ongoing Extensions:
  – Diverse & Partial Satisfaction Planning
  – BURLAP
  – HUBO Robotic Platform (Constraint-based Planning)
ActorSim: The Actor Simulator

Goal Refinement Library
- LifeCycle
- Memory

ActorSim Core
- Common Abstractions

ActorSim Planner

ActorSim Connector
- Coordination Executive
- Team Executive(s)
- Domain Knowledge

Third Party Executives & Simulators
- MASON
- GRIM/MAGR
- Minecraft
- StarCraft
- Roomba
- ROS/Gazebo
- TBD

http://makro.ink/actorsim
ActorSim: The Actor Simulator

- Goal Refinement Library
  - LifeCycle
  - Memory

- ActorSim Core
  - Common Abstractions

- ActorSim Planner

- ActorSim Connector
  - Coordination Executive
  - Team Executive(s)
  - Domain Knowledge

- Third Party Planners

- Third Party Executives & Simulators
  - MASON
  - GRIM/MAGR
  - Minecraft
  - StarCraft
  - Roomba
  - ROS/Gazebo
  - TBD

http://makro.ink/actorsim
Online Planning

• State Transition System $\Sigma = (S, A, E, \gamma)$
  – **State**: $s \in S$
  – **Action**: $a \in A$
  – **Event**: $e \in E$
  – **Transition Function**: $\gamma: S \times (2^A \cup E) \rightarrow 2^S$

Goals are often static and provided

• Planner: $M_\Sigma \times S \times G \rightarrow \pi$
  – **Model of $\Sigma$**: $M_\Sigma$
  – **Goal**: $g \in G \subseteq S$
  – **Plan**: $\pi = [a_1, a_2, ..., a_n] \in \Pi$

• Controller: $S \times G \times \Pi \rightarrow S \times \Pi$

Nau (2007)
Technical Approach: Goal Reasoning

Goal Reasoner (GRPROCESS) manages and revises goals and their priorities

Goal Memory (M)

Sets $S_g$ or adjusts $\Sigma$

Descriptions of $\Sigma$, $s_0$, and $S_g$

Planner

Sets

Scheduler

Expands $\Pi$ or replans

Controller

Commits to $\pi \in \Pi$ or repairs

System $\Sigma$

Goals

Execution Status

Observations

Actions

Events

Goal-Task Networks (Alford, Shivashankar, Roberts, Frank, Aha, IJCAI-16)

- Hybrid Planning formalism
- Blends Hierarchical Goal Networks and Hierarchical Task Networks

Nau (2007)
Goal Reasoning in a Nutshell

Creation

Intention

Planning & Scheduling

Execution Monitoring

Goal Memory

Goal Lifecycle

FORMULATE

FORMULATED

REFORM

SELECT

SELECTED

DEFER

EXPAND

EXPANDED

REPLAN

COMMIT

COMMITTED

REPAIR

DISPATCH

ATTACHED

CONTINUE

EVALUATE

EVALUATED

RESOLVE-BY

Monitor
Goal Reasoning Theory

- Goal Reasoning Theory (*Roberts, et al., 2016, ACS*)
  - Extends Goal-Task Networks to Goal reasoning
  - Provides a clear semantics for GR processes
  - Formalizes semantics of ActorSim’s implementation

Goal State Transition System \( Z = (M, R, \gamma_{GR}) \)

\( M \) is a goal memory where \( M_{ij} \)
- \( g_i \) is the \( i^{th} \) goal \( (0 < i < m) \)
- \( m_j \) is the \( j^{th} \) metric \( (0 < j < n) \)

\( R \) is a set of **refinement operators**

\( \gamma_{GR} : M \times R \rightarrow M' \)
ActorSim: The Actor Simulator

http://makro.ink/actorsim
Goal Reasoning: The Goal Lifecycle

Diagram showing the goal lifecycle stages:
- Formulate
- Formulated
- Select
- Selected
- Expand
- Expanded
- Commit
- Committed
- Dispatch
- Dispatched
- Evaluate
- Evaluated
- Monitor
- Adjust
- Drop
- Finish
- Process

Stages with feedback loops:
- Reform
- Resolve-to
- Defer
- Resolve-to
- Replan
- Resolve-to
- Repair
- Resolve-to
- Continue
- Resolve-to
- Resolve-by
The Goal Lifecycle Without Error Strategies
The Goal Lifecycle For Replanning

1. **FORMULATE**
   - FORMULATED
2. **SELECT**
   - SELECTED
3. **EXPAND**
   - EXPANDED
4. **COMMIT**
   - COMMITTED
5. **DISPATCH**
   - DISPATCHED
6. **EVALUATE**
   - EVALUATED
7. **RESOLVE-BY**
8. **PROCESS**
9. **FINISH**
10. **DROP**
Alex’s Goal-Task Network

Key
- Goal
- Task
- Action

- Complete Obstacle Course
- Gather Resources
- Survive

Move To Target

- StepAroundTo
- StepTo
- StairsTo
- BridgeTo
- MineTo

Move One Block

Place Block

Mine Block

Craft Item

Eat

Hit Entity

Craft Tool

Mine Resource

Stay Full

Defend
Selecting Subgoals using Deep Learning

Selecting Subgoals using Deep Learning

**Survive**
- Complete Obstacle Course
- Complete Course
- Achieve at(target)
- Achieve next-closest
  - Complete Course
  - Create Course
  - Move To Target
  - Move One Block
  - Place Block
  - Mine Block
- Gather Resources
- Craft Tool
- Mine Resource
- Craft Item
- Eat
- Hit Entity
- Stay Full
- Defend
- Move To Target
- StepAroundTo
- StepTo
- StairsTo
- BridgeTo
- MineTo
- Survive
- Stay Full
- Defend

- Move One Block
- Place Block
- Mine Block
Goal Reasoning Walkthrough

- Complete Course
  - Create Course
    - Achieve at (target)
  - Achieve next-closest

Diagram:

1. **FORMULATE**
   - FORMULATED
2. **SELECT**
   - SELECTED
3. **EXPAND**
   - EXPANDED
4. **COMMIT**
   - COMMITTED
5. **DISPATCH**
   - DISPATCHED
6. **EVALUATE**
   - EVALUATED

Processes:
- PROCESS
- FINISH
- DROP

Resolve-by:
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)
- Complete Course
  - Achieve at(target)

Unformulated Expansions

Diagram:
- FORMULATE
  - FORMULATED
  - SELECT
  - SELECTED
  - EXPAND
  - EXPANDED
  - COMMIT
  - COMMITTED
  - DISPATCH
  - DISPATCHED
  - EVALUATE
  - EVALUATED
  - RESOLVE-BY
  - PROCESS
  - FINISH
  - DROP
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

Complete Course

Achieve at(target)

Achieve next-closest

Unformulated Expansions

- FORMULATE
  - FORMULATED
- SELECT
  - SELECTED
- EXPAND
  - EXPANDED
- COMMIT
  - COMMITTED
- DISPATCH
  - DISPATCHED
- EVALUATE
  - EVALUATED
- RESOLVE-BY
- PROCESS
- FINISH
- DROP
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

- Complete Course
- Achieve at(target)
- Achieve next-closest

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

- Complete Course
- Achieve at(target)
- Achieve next-closest

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

Complete Course

Achieve at(target)

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

- Complete Course
- Achieve at(target)
- Achieve next-closest

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)
- Complete Course
- Achieve at(target)
- Achieve next-closest

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

- Complete Course
- Achieve at (target)
- Achieve next-closest

Unformulated Expansions

FORMULATE

FORMULATED

SELECT

SELECTED

EXPAND

EXPANDED

COMMIT

COMMITTED

DISPATCH

EVALUATE

EVALUATED

RESOLVE-BY

FINISH

DROP

PROCESS
Goal Memory (formulated goals)

Complete Course

Achieve at(target)

Goal Reasoning Walkthrough

Unformulated Expansions
Goal Reasoning Walkthrough

**Goal Memory (formulated goals)**
- Complete Course
- Achieve at(target)

**Unformulated Expansions**
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

Unformulated Expansions
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

- Complete Course

Unformulated Expansions

FORMULATE

FORMULATED

SELECT

SELECTED

EXPAND

EXPANDED

REPLAN

COMMIT

COMMITTED

DISPATCH

DISPATCHED

EVALUATE

EVALUATED

RESOLVE-BY

PROCESS

FINISH

DROP
Goal Reasoning Walkthrough

Goal Memory (formulated goals)

Unformulated Expansions
Video Removed to decrease disk size
ActorSim: The Actor Simulator

- Goal Refinement Library
  - LifeCycle
  - Memory
- ActorSim Core
  - Common Abstractions
- ActorSim Planner
- ActorSim Connector
  - Coordination Executive
  - Team Executive(s)
  - Domain Knowledge
- Third Party Executives & Simulators
  - MASON
  - GRIM/MAGR
  - Minecraft
  - StarCraft
  - Roomba
  - ROS/Gazebo
  - TBD

Experimental Setup

- **Obstacle Course**
  - Lava, Pond, Short/Tall Wall, Pillar, Empty, Stairs, Arch, Comb

- **Capture Expert Traces (Upper Bound)**
  - Four actions: Step, Stairs, Bridge, Mine
  - Observe local state to choose best subgoal
  - Collect (screenshot, action-chosen) pairs

- **Deep Learning of Classification Task**
  - AlexNet Architecture
  - Lowered learning rates for 7 layers
  - Replaced final layer
  - Learned using FlickrStyle solver weights & iterations
Data Capture

• Screen capture from behind agent
• 10 training courses with random sections (~13K frames)
  – 9 courses of 100 sections from 9 angles as in figure
  – 1 course of 500 random sections manipulated angles
• 1 testing course (892 useable frames)
  – 100 sections with manipulated angles
Training & Testing

• Equalized instances to 348 frames in 4 action classes
  – 80% (278) selected from each class for training
  – 20% (70) selected from each class for validation
  – Training and validation data shuffled

• Training Cost
  – Trained on a Tesla K40
  – Completed 1M iterations in less than 10 hours

• Testing data from final course run
Results

- Achieved 92% accuracy on training data
- Quickly reached best values within an hour
- Achieved 87% (777/892) on testing data
Ongoing research in Deep Learning

• Challenge: occlusion of blocks by character
  – Add recursive structure to network
• Challenge: reward signal for reinforcement learning
  – Obtain reward from high-level goals
• Challenge: maintaining separate policies for tasks
  – Apply hierarchical approaches
• Challenge: current network is separate from ActorSim
  – Integrate learned network into ActorSim
Ongoing research in ActorSim

Challenge in long-life actors: How to focus learning effort?

A perpetual learner directs its own curricula to continually:
- learn new tasks,
- revise known tasks, or
- halt learning on tasks it has already mastered.

• Automated Planning
  – Diverse Planning (with M. Floyd)
  – Partial Satisfaction Planning (with P. Bevan)

• Reinforcement Learning
  – BURLAP (MacGlashan et al.)

• Foreign Disaster Relief (Johnson et al., 2016, ACS)
  – Goal Reasoning with Information Measures
ActorSim Toolkit

- **ActorSim**: Minecraft
- **Goal Reasoning Theory**
- **Future Directions**
  - GTN Planning
  - Deep Learning
  - Reinforcement Learning

Video Removed to decrease disk size

More Details:
Dave Bonanno or Mark “Mak” Roberts
ActorSim: http://makro.ink/actorsim
{david.bonanno,mark.roberts.ctr}@nrl.navy.mil