Performance Measures, Environments, Actuators, Sensors (PEAS) for Testing Autonomous Intelligent Agents

Dr. Laura Freeman, Dr. Jitesh Panchal
Adam Dachowicz, Melanie Grande, Andrew Lang (Purdue)
Dr. Ivan Hernandez, Scott Welch (Virginia Tech)

12th Annual SERC Sponsor Research Review
HOSTED VIRTUALLY ON: November 18, 2020
www.sercuarc.org
Outline

• Phase 1: Test and Evaluation Framework

• Phase 2: Performance Measures Case Studies

• Summary & Future Work
Test and Evaluation Framework

- Multi-Agent Systems of Autonomous Intelligent Agents
VTP Model
Hierarchy of Test

- Digital Simulation
- Software in the loop
- Hardware in the loop
- Ground testing
- Deployment to Space

Increasing fidelity

Increasing test points
Framework Process

- Phase
  - Field of Study
  - Hierarchy of Test
- Goal of Test
- Test Plan Efficiency
- Test Methods
- Test Plan
## Execution Illustration

<table>
<thead>
<tr>
<th>System Under Test</th>
<th>Phase</th>
<th>Field of Study</th>
<th>Hierarchy of Test</th>
<th>Test Plan Efficiency</th>
<th>Goal of Test</th>
<th>Test Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI Algorithm</td>
<td>Unit Testing</td>
<td>AI Mission Environment</td>
<td>Components: Functions in code, Program integrates functions, Simulated inputs</td>
<td>- Blackbox testing as scripts on simulated program inputs, Focus on bugs or misunderstood requirements to stress all possible types of inputs and conditions</td>
<td>Assess AI reliability, robustness, performance, and biases</td>
<td>CIT combined with optimal learning to ensure input space coverage and minimize unknown areas of performance</td>
</tr>
</tbody>
</table>

Table shows a high-level overview of how each framework concept contributes to guiding test plan development.
## Execution Illustration

<table>
<thead>
<tr>
<th>System Under Test</th>
<th>Phase</th>
<th>Field of Study</th>
<th>Hierarchy of Test</th>
<th>Test Plan Efficiency</th>
<th>Goal of Test</th>
<th>Test Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite Systems</td>
<td>Integration Testing</td>
<td>Above Plus: - Software - Electronic Hardware - Mechanical Systems</td>
<td>Components: - AI Algorithm, Sensors, Actuators, Control Software - Integration considers inputs from/outputs to components to other subsystems</td>
<td>- Integration testing of full system in simulated space environment - Focus on coverage of interactions between systems to identify faults or performance drop-offs</td>
<td>Assess system performance, reliability, cybersecurity, and interoperability</td>
<td>DOE using both known software integration areas and environment factors</td>
</tr>
<tr>
<td>System Under Test</td>
<td>Phase</td>
<td>Field of Study</td>
<td>Hierarchy of Test</td>
<td>Test Plan Efficiency</td>
<td>Goal of Test</td>
<td>Test Methods</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>--------</td>
<td>-----------------------------------</td>
<td>-----------------------------</td>
<td>----------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Constellation with Ground Station Operators</td>
<td>Operation</td>
<td>Above plus: - Psychology - Human-Computer Interaction - AIA teaming</td>
<td>Components: - Satellites - Ground station - Operators - Environmental conditions</td>
<td>- System in operational environment - Focus on achieving mission objectives, detecting changes in performance</td>
<td>Assess mission accomplishment</td>
<td>DOE with human's as factors (within-subject design)</td>
</tr>
</tbody>
</table>
Outline

- Phase 1: Test and Evaluation Framework
- Phase 2: Performance Measures Case Studies
- Summary & Future Work
## Case Study of System of Cognitive Agents

### Mission Tasking

<table>
<thead>
<tr>
<th>Mission</th>
<th>System-Level Task (Goal)</th>
<th>Responsible Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Observation plan (locations, orientations, and time) created</td>
<td>Flight Controller</td>
<td></td>
</tr>
<tr>
<td>2. Task schedule created for all spacecraft</td>
<td>Flight Controller</td>
<td></td>
</tr>
<tr>
<td>3. Data packed and sent to appropriate agents/stations</td>
<td>Flight Controller</td>
<td></td>
</tr>
<tr>
<td>4. Data and commands received and unpacked</td>
<td>Flight Controller</td>
<td></td>
</tr>
<tr>
<td>5. Spacecraft in position at next observation location</td>
<td>Movement Agent</td>
<td></td>
</tr>
<tr>
<td>6. Observation data taken of desired location</td>
<td>Mission Manager</td>
<td></td>
</tr>
<tr>
<td>7. Fires recognized and characterized</td>
<td>Processing Agent</td>
<td></td>
</tr>
<tr>
<td>8. Subsystems operating to accomplish their goals, with all faults responded to</td>
<td>Recovery Agent</td>
<td></td>
</tr>
</tbody>
</table>

### Observation Plan

- **Location**: Forest fires in southwestern United States during summer
- **Purpose**: To detect and characterize forest fires
- **Process**:
  1. **Communication of status and data** from Flight Controller to Mission Manager and other agents.
  2. **Mission Tasking** involving coordination and task negotiation.
  3. **Flight Controller** creates the observation plan (locations, orientations, and time) for spacecraft.
  4. **Flight Controller** schedules tasks for all spacecraft.
  5. **Data packed and sent to appropriate agents/stations**.
  6. **Data and commands received and unpacked** by Flight Controller.
  7. **Spacecraft** move to the next observation location.
  8. **Observation data** taken of desired location.
  9. **Agents** recognize and characterize fires.
  10. **Subsystems operating** to accomplish their goals, with all faults responded to.

### System Flow

- **Team of Agents (1)**
  - Movement Agent
  - Flight Controller
  - Recovery Agent
  - Processing Agent

- **Team of Agents (2)**
  - Flight Controller 1
  - Flight Controller 2

- **Communication** of status and data from Flight Controller to other agents/stations.
“Onion” Model of Performance Measures

- **Black box** evaluation examines observable outcomes
- **White box** evaluation examines the implementation algorithms
- **Grey box** evaluation represents a place along the continuum where there some access to the system’s processes

Mission Objectives:
- Provide canopy coverage over the United States
- Detect and track fire plumes in the United States
**Task:** The work a system will undertake to achieve the mission.

A **Task** must be comprised of 3 parts:

1. **Goal:** a desired discrete end state
2. **Objective:** a criterion describing how well the goal was achieved
3. **Constraint:** a restriction or condition placed on the variables needed to achieve the goal

**Sub-Task:** A task necessary to complete prior to completion of the higher-level task.

**Method:** Can be algorithms, models, etc. that are commonly successful at performing the task.

**Data Collected:** Data or measurements that need to be collected, usually required to compute some performance measure at the end.
### Summary of Results

#### “Onion” Model of Performance Measures

#### Measures of Effectiveness
- Mission Success - % Capture
- Canopy Coverage Completeness
- Fire Detection - Precision
- Fire Detection - Recall
- Probability of Tipping Success
- Fire Track Completeness
- System Availability
- System Recovery Timeliness

<table>
<thead>
<tr>
<th>Agent</th>
<th>Agent Level PM</th>
<th>Team Level PM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flight Controller Agent (part 1 – tasking agent)</td>
<td>Degree of satisfaction of preference constraints; utilization efficiency measure</td>
<td>Aggregate utilization efficiency</td>
</tr>
<tr>
<td>Flight Controller Agent (part 2 – communications agent)</td>
<td>Error rate (of command data being unpacked)</td>
<td>Aggregate data sets transferred to ground control over time</td>
</tr>
<tr>
<td>Flight Controller Agent (part 3 – observation agent)</td>
<td>Percentage of ground observed</td>
<td>Aggregate revisit time, aggregate number of fires and properties identified, number of visits per location (or length of tracking time)</td>
</tr>
<tr>
<td>Mission Manager Agent</td>
<td>Duty cycle of sensors, error (MAE/RMSE) of predicted vs. actual power draw, error (MAE/RMSE) of area observed vs. desired area observed</td>
<td>N/A</td>
</tr>
<tr>
<td>Spacecraft Movement Agent</td>
<td>Propellant usage, reaction wheel saturation, error (MAE/RMSE) of predicted vs. actual propellant usage</td>
<td>N/A</td>
</tr>
<tr>
<td>Spacecraft Processing Agent</td>
<td>Number of true/false positives, number of true/false negatives, number of fires and properties identified</td>
<td>Aggregate number of fires and properties identified</td>
</tr>
<tr>
<td>Spacecraft Recovery Agent</td>
<td>Downtime of each system over time frame</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Outline

• Phase 1: Test and Evaluation Framework

• Phase 2: Performance Measures Case Studies

• Summary & Future Work
Summary

• This year we:
  – Developed an overarching framework for thinking about test and evaluation
  – Developed case study specific performance measures for mission/tasks/agents
  – Initiated research into Agent-Team performance measures and analysis methods that would enable performance prediction in previously un-tested environments/tasks

• Deliverables
  – T&E Framework Report
  – Performance Measures for Multi-Agent Systems of Autonomous Intelligent Agents in Satellite Systems
    o Case Study Report
• Test and Evaluation Program should:
  – Establish a method for collecting data
    o This data should include performance measures useful for evaluation and prediction and other descriptive data that characterize the mission, tasks, and operating environment and provide diagnostic capabilities for explaining changes in performance measures
  – Establish a method for determining sufficient data collection to meet algorithm assurance requirements
  – Be useful for performing all the following decision-making functions:
    o Guide system and algorithm design by collecting data on the evolution of system capabilities, allowing for the iteration of multiple system architectures and algorithm selections
    o Understand system performance growth (or degradation) as a function of system maturity
    o Characterize performance as a function of operating environments and tasks
    o Make predictions about future environments/tasks
    o Quantify risk and make informed decisions based on risk reduction or risk tolerance
Future Work Vision

- Expect there is a high priority in getting fielded capabilities into space over the next 5 years
- Test and evaluation is a fundamental capability required to make informed decisions on fielding and operational use
- Future work should address both the direct application to real systems and expanded investigation into what new measurement and analysis capabilities are needed
- Proposed three strategic areas:
  - Implemented and integration of performance measures with programs
  - Expand and refinement of performance measures for:
    - Agent Teams
    - Cognitive architectures – expand performance measure development with SOAR architecture specific measures
  - Close the “Design – Test – Evaluation – Design” cycle
    - Supporting design of cognitive agents via trade studies of engineering configurations
    - Develop an online test and evaluation