Wrap-Up: Metrics and Evaluation

Search/Coverage, Part I

January 30, 2003

Class Meeting 6
Objectives

• Understand evaluation techniques for multi-robot systems
  – Theoretical proofs
  – Empirical studies

• Paper #7 presentation: Chris Reardon

• Search/Coverage, Part I
Techniques for Evaluating Robot Control Code

• Theoretical / Analytical:
  – Proofs of correctness

• Empirical:
  – Multiple experimental runs
  – Experiments to eliminate effect of specific starting conditions, environments, parameter settings, etc.
  – Average experiments over many trials
  – Maintain data on averages and standard deviations
  – Number of experiments needed: enough to ensure statistical significance
Key Points of Theoretical / Analytical Evaluation

• Proofs of correctness: can be used to theoretically validate multi-robot control code.

• However, in general, it is difficult to generate such proofs.

• Many assumptions must be made; some assumptions may not be true in robots’ physical world.

• Therefore, it’s more common to validate multi-robot control code empirically (i.e., experimentally)
Empirical Evaluation and Validation of Multi-Robot Control Code

- **Main point:** since it is difficult to sufficiently model robot team’s environment and interactions, it’s better to validate multi-robot control code empirically.

- **However,** in must ensure that performance evaluation is not skewed through limited test case selection.

- **Must generalize over:**
  - Range of parameter settings
  - Range of robot environments (or, at least across a certain class)
  - Numbers of robots
  - Etc.
Typical Experimental Evaluation Approach

• Define important variables/parameters of the system
• Determine a range of values for variables/parameters to be tested
• Define metrics to be used for evaluating system
• Determine characteristics of environment which algorithm is designed to handle
• Generate a variety of environments within this class to test
  – Best if environments are randomly selected
  – For example, generate obstacle size and location based upon a particular distribution
• Determine a baseline performance to which the new algorithm will be compared
• Run multiple experiments, varying starting positions of robots randomly
• Collect data over multiple experimental runs, and average data, also maintaining standard deviation
• Determine how many experimental runs are needed to generate statistical significance
• Continue data collection until results are significant, or no significant difference is found
Calculating Statistical Significance

• Collect data from 2 or more cases according to selected metric
• Compute average and standard deviation
• Use Student’s t distribution, comparing policies 2 at a time
• In these computations, use null hypothesis:
  – $H_0$: $\mu_1 = \mu_2$, and there is essentially no different between the two policies.
• Under hypothesis $H_0$:

$$ T = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} $$

where:

$$ \sigma = \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}} $$
Calculating Statistical Significance (con’t.)

• Then, on the basis of a two-tailed test at a 0.01 level of significance, we would reject $H_0$ if $T$ were outside the range $-t_{.995}$ to $t_{.995}$

• Example: For $n_1 + n_2 - 2 = 250 + 250 - 2 = 498$ degrees of freedom,
  $-t_{.995}$ to $t_{.995}$ is in the range $-2.58$ to $2.58$.
  (You can find these values in probability and statistics textbooks.)

• If $T$ value is outside this range, we can reject $H_0$ at a 0.01 level of significance

• If $H_0$ rejected, can conclude that the variation in performance of 2 policies is statistically significant.
Thought Exercise: Clearing Land Mines

- Fill in the blank:
  - Important variables/parameters of system
  - Range of values of parameters to be tested
  - Metrics
  - Characteristics of environment
  - How to generate test environments?
  - Baseline algorithm
  - How many experiments to run?
  - How to determine statistical significance of data?
“Benchmark Missions” for Robot Team Evaluation

• Multi-Robot Soccer

• Search and Rescue – NIST Arena
Summary of Multi-Robot Control Code Evaluation

• Two primary approaches to evaluating control code:
  – Theoretical / analytical
  – Empirical
• Theoretical / analytical based upon proofs of correctness or fundamental limitations of approaches
• However, theoretical / analytical techniques can be difficult to develop due to need for unrealistic assumptions, difficulty in modeling, etc.
• Empirical approaches more common in robot control evaluation
• Must ensure that empirical data is relevant through:
  – Multiple runs varying parameters of interest
  – Randomly generating starting conditions
  – Statistically significant data

Presented by Chris Reardon
The Search/Coverage Problem and Sensor Networks

• For the next 3 classes – focus on distribution of multiple homogeneous robots/sensors for purpose of detecting targets / items of interest

• Large literature on “Search” in operations research community

• We’ll look at a sampling of this work as applied to distributed robotics
Example of Sensor Net Coverage Implementation

(From A. Howard’s work, USC)
The Coverage Problem:

- Application of sensor or effector to an extended space

Potential applications include:

- Floor cleaning
- Characterization
- Mine sweeping
- Reconnaissance
- Sentry duty
- Communications relay
- Maintenance inspection
- Carrier deck FOD disposal
- Ship hull cleaning
Issues Studied in Optimal Search

- Target Distribution
- Detection Models
- Uniformly Optimal Search Plans
- Properties of Optimal Search Plans
- Search with Discrete Effort
- Optimal Search and Stop
- Search in Presence of False Targets
- Optimal Search for Stationary Targets
- Approximation of Optimal Plans
- Conditionally Deterministic Target Motion
- Markovian Target Motion
- Etc.

We’ll just skim the surface!
Gage’s Search/Coverage Research: Agents/Robots of Coverage System

• Large enough number of agents that human operator can’t deal with each agent individually

• Each element possesses:
  – Some amount of mobility
  – One or more mission-capable sensors or effectors
  – (Optionally) A navigational sensor capability enabling relative or global positioning
  – (Optionally) A communications capability
  – Processing capability

Much of this discussion derived from Gage’s work. See:
Types of Coverage

• Three types of coverage:
  – Blanket/Field (e.g. minesweeping)
  – Barrier (e.g. intrusion prevention)
  – Sweep (e.g. foreign object debris (FOD) removal)

• Optimal coverage: desired ensemble behavior is maintenance of a spatial relationship between agents that is appropriate to the mission, and which adapts to specific local conditions.

• Note different goals for search/coverage:
  – Maximize target detections per amount of search effort (e.g., prospecting)
  – Minimize number of targets missed in a swept area (e.g., de-mining)
  – Minimize expected time for target detection (e.g., lost child)
Blanket/Field Coverage

- **Objective:** to achieve a static arrangement of elements that maximizes the detection rate of targets appearing within the coverage area.
Barrier Coverage

- **Objective:** achieve a static arrangement of elements minimizing the probability of undetected penetration through the barrier.
Objective: move elements across a coverage area in a way that balances:

- Maximization of number of detections per time
- Minimization of the number of missed detections per area.
Also Need to Address Boundary Conditions

• Coverage behaviors: “Steady state”

• “Boundary conditions”:
  – Deployment
  – Recovery
  – Navigation of the group as a whole

• Other factors to be considered include:
  – Obstacles and traversability
  – Randomness of behavior
  – Analyzing internal dynamics of system using analogies from biology:
    • Herding, schooling, immune system and pheromone mechanisms
    and physics:
      • Entropy, temperature, pressure, solid, liquid, gas
Preview of Next Class

• Search/Coverage, Part II

• Paper Presentation: Yifan Tang