Final Project Proposal: Mission planning and exploration via POMDP

Stuart Johnson stujohn@stanford.edu

AA228/CS238, Stanford University

1. Goal

An autonomous agent needs to locate and carefully inspect (by photography, say) a series of electronic devices which have been deployed to a geographic region. The agent must therefore find these devices and proceed to their locations. These devices were deployed in such a way that their locations are only approximately known - that is to say, we have priors on their locations. In order to find these devices, the agent can use radio direction and ranging equipment to communicate with the devices - which are passively listening and capable of responding (one at a time) when interrogated by radio. This sensing returns a noisy relative location of the device - but the agent is capable of improving the location estimate as measurements are made. Both the agent and the devices require power to perform these communications - the devices are powered, but minimally. The agent has more power, but the radio direction finding is a power-hungry operation. Once within a given range of the object, the agent can see the device, and can proceed directly to complete its inspection task. If either the agent or the devices exceeds power consumption in a given time frame, they must cease operations to recharge (one imagines solar power). The recharge time is known. The agent also requires power to move.

The goal is to find a movement and sensing schedule (a policy) for the agent to maximize the utility of a mission given a fixed mission time. It will be more important to inspect some devices than others - for example some may be more mission-critical, so that the utility will depend on the device. Also, the interplay of the various parameters of the device and agent and the optimal policy are of interest to gain insights into system design.

This problem was inspired by interest in posing the sport of orienteering (but NOT "team orienteering") as a POMDP. That project is still in discussion with an orienteer. This problem is reminiscent of SLAM as a POMDP (Placed et al., 2023), but our goals are not to make a better map - we have a specific mission in mind. If we were merely interested in improving our estimates of the device positions, this is a type of resource-constrained exploration problem (Peltzer et al., 2022). Such problems are ubiquitous.

2. Decision Making

The route (target order selection and travel) and radio communication actions are all sequential decisions. The quality of the location of the devices is an integral of the running communication efforts and dependent on all past actions. The sequence of actions results in distinct branches in the utility of a given policy. For example, early attempts to communicate with many devices (or perhaps even worse - just one!) in order to improve the map quality may cause power issues later in the mission.

This problem maps to the following POMDP.

- States:

  - agent: position, velocity and heading, battery level, operational state
  - each device: position, battery level, operational state
- Actions:
  - agent: ping device i (i.e. acquire device bearing and distance)
  - agent: set agent heading and velocity
- Transitions:
  - agent: execute dynamics/kinematics (deterministic)
  - each device: enter or exit recharge
  - agent: enter or exit recharge
  - agent/device: communication/localization
- Rewards:
  - device: communication energy penalty (negative)
  - agent: communication/localization energy penalty (negative)
  - agent: movement energy penalty (negative)
  - agent: time penalty
  - agent: recharge penalty (optional)
  - agent: device specific inspection reward (the only reward!)
- Observations:
  - device i heading and range (relative to agent)

3. Sources of Uncertainty

The location of each device is uncertain. This results from location priors and the error in the heading and range acquired by the agent from a device ping. Knowledge of the locations of the devices will improve as the mission progresses.

4. Sketches of Solution (Optional)

The solution might include such diverse elements as:

- parameterization: A 2D Grid world with simplified dynamics (up,down,left,right). Grid resolution would presumably need to be sufficient not to cause significant artifacts. A more efficient graph discretization of space might be better for problems with route constraints, or if we can argue for it as a sufficient representation of movement space.
- Belief approximations (e.g. (Kochenderfer et al., 2022), Chapter 19)
- policy search: We do not necessarily need the globally optimal solution - a workable policy to confirm system design is sufficient - so the solution is open to approximate methods. It is not clear to me how much approximation will be necessary.

References

Mykel J. Kochenderfer, Tim A. Wheeler, and Kyle H. Wray. *Algorithms for Decision Making*. MIT Press, 2022.

Oriana Peltzer, Amanda Bouman, Sung-Kyun Kim, Ransalu Senanayake, Joshua Ott, Harrison Delecki, Mamoru Sobue, Mykel Kochenderfer, Mac Schwager, Joel Burdick, and Ali akbar Agha-mohammadi. Fig-op: Exploring large-scale unknown environments on a fixed time budget, 2022. URL https://arxiv.org/abs/2203.06316.

Julio A. Placed, Jared Strader, Henry Carrillo, Nikolay Atanasov, Vadim Indelman, Luca Carlone, and José A. Castellanos. A survey on active simultaneous localization and mapping: State of the art and new frontiers, 2023. URL https://arxiv.org/abs/2207.00254.