Multi-Agent Planning under Uncertainty with Rare Catastrophic Events

Youngjun Kim
Department of Aeronautics and Astronautics
Stanford University
496 Lomita Mall
Stanford, CA 94305

Abstract

This dissertation abstract outlines some of theoretical frameworks for modeling and simulation of multi-agent planning problems with rare catastrophic events. In particular, this abstract will focus on a wildfire surveillance application using unmanned aircraft. The thesis abstract presents an initial model and results of a preliminary study.

Introduction

Optimal planning in problems with multiple agents interacting in an uncertain environment is extremely challenging. Many problems are computationally intractable because of its large state and action spaces. In addition, some problems involve rare catastrophic events that significantly influence expected utility. The rarity of events can make computing the optimal policy challenging. The thesis will propose efficient methods for solving problems with multiple agents interacting in the presence of rare catastrophic events, with application of wildfire surveillance using unmanned aerial vehicles (UAVs). The thesis abstract briefly explains the wildfire surveillance problem, its challenges, and possible approaches for solving the problem. It shows results of a preliminary study and concludes with future works. Due to the complexity of modeling the problem and methods to solve the problem, the thesis abstract introduces a research plan that includes defining the scenario, applying methods of rare event simulation, modeling sequential decisions of an agent, modeling an agent as human, and finally formulating a two-players game with rare events.

Wildfire Surveillance using UAVs

Motivation

In 2013, the third largest wildfire in California's history started in the Sierra Nevada mountain range but soon reached Yosemite National park. It burned more than 200,000 acres for nine weeks. Because the area of the wildfire was geographically distributed, it was hard to monitor the area without aerial support. The MQ-1 Predator UAV was introduced for surveilling the area.

Unmanned aircraft have been occasionally used for surveilling large wildfires since 2007 to capture infrared images of Southern California fires. There is tremendous interest in the use of unmanned aircraft for wildfire surveillance because of their low operational cost, ability to operate in harsh weather conditions, and reducing risk to pilots and firefighters. The images taken from unmanned aircraft can help improve the ability of authorities to predict the evolution of fires and support decision making of an incident commander about where to allocate suppression resources.

Challenges

Although there are many benefits to using unmanned aircraft, there are some issues when integrating unmanned aircraft into current wildfire surveillance operations. One of the biggest concerns is the risk of collision with manned aerial supports and other unmanned aircraft. Firefighters in the field want to launch low cost, hand-held aircraft for obtaining surveillance information. The airspace will be shared by low-altitude unmanned aircraft and manned aerial supports such as helicopters, spotters, or tankers. It is important to maximize surveillance but without compromising safety.

An important issue when unmanned aircraft are operated in the field is communication loss between the aircraft and its pilots on the ground. More than 400 large U.S. military drones have crashed around the world since 2001 (Whitlock 2014). Unreliable communication links is one of the primary reasons for vehicle loss. Unmanned aircraft operated in the vicinity of wildfires would be under unfavorable conditions for maintaining reliable radio communication. Communication loss needs to be considered when planning the actions of unmanned aircraft.

In addition, plans developed for unmanned aircraft monitoring wildfires should account for the various sources of uncertainty. Fires propagate stochastically based on environmental factors such as weather and topological conditions. Manned aircraft in the vicinity follow might follow preplanned paths, but but they may change course based on how the wildfire propagates. Unmanned aircraft should consider the uncertainty of the manned aircraft route to avoid collision. Adding to the uncertainty of the manned aircraft route, observation of locations of the unmanned aircraft and manned aircraft can be noisy and incomplete.

Approach

This section outlines possible approaches for modeling and simulation of the wildfire surveillance problem.

Rare Event Simulation

Rare event simulation is especially challenging and inherently requires heavy computational effort due to the rareness of the event. Aircraft collision risk estimation (Kim and Kochenderfer 2015) is one example of rare event simulation. Estimates of mid-air collision risk can be obtained through Monte Carlo simulation of encounters sampled from a probabilistic airspace model (Kochenderfer et al. 2008; 2010). Due to the rarity of collision events, typically millions of simulations are required. Techniques known as importance sampling and the cross-entropy method (De Boer et al. 2005; Rubinstein and Kroese 2004) have been used in the past to bias the sampling on trajectories that are likely to result in collision. Reliable estimates of collision risk can be obtained with only a fraction of the computational cost required by crude Monte Carlo simulation.

These techniques can be applied to solving a single-shot decision problem presented later in the abstract. Since collisions between unmanned aircraft and manned aircraft are rare events, these techniques can help improve the speed of simulations.

Direct sampling If X is a discrete random variable and its probability mass function is $f_{\mathrm{X}}(x)$, the expected value of a function g of X is shown in Eq 1.

$$\mathrm{E}(g(\mathrm{X})) = \sum_{x \in \mathrm{X}} g(x) f_{\mathrm{X}}(x) \tag{1}$$

Eq 2 is an unbiased estimator of the expected value. It takes n samples, $(x_1, \cdots, x_n)$, and computes the mean of $g(x)$ over the samples.

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i) \tag{2}$$

If we want to estimate the probability of a rare event and $g(x)$ indicates whether x is a rare event, most of samples are not useful and the estimator requires a lot of samples to have a reasonable estimate.

Importance sampling Importance sampling is choosing a good distribution from which to simulate a random variable X. Samples are drawn from a proposal distribution h_{X} that generates more rare events instead of sampling from the original distribution f_{X}, and samples are weighted properly as shown in Eq 3

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i) \frac{f_{\mathrm{X}}(x_i)}{h_{\mathrm{X}}(x_i)} \tag{3}$$

With a good proposal distribution, a better estimate can be computed with fewer samples than direct sampling.

Cross-Entropy method The cross-entropy method provides a systematic way to find a good proposal distribution for importance sampling. It is an adaptive algorithm involving an iterative procedure. Each iteration is broken down into two phases:

1. Generate random samples from a proposal distribution
2. Update the parameters of the proposal distribution based on the samples to produce better samples in the next iteration.

Sequential Decision under Uncertainty

There are many uncertainties in this problem. It is uncertain when communication is lost and when it returns. Observation about the location of manned aircraft is noisy. The problem will be formulated as a partially observable Markov decision process (POMDP) (Kochenderfer 2015). A POMDP is often used to formulate a sequential decision problem with state uncertainty. It tracks the belief about the current state. There are few simple offline methods to solve POMDP problems such as QMDP and Fast Informed Bound (FIB) (Hauskrecht 2000). Offline approximate POMDP solution algorithms have focused on point-based approximation techniques, as surveyed by (Shani, Pineau, and Kaplow 2013), such as Point-Based Value Iteration (PBVI) (Pineau, Gordon, and Thrun 2006), Heuristic Search Value Iteration (HSVI) (Smith and Simmons 2004), and Successive Approximations of the Reachable Space under Optimal Policies (SARSOP) (Kurniawati, Hsu, and Lee 2008). However, these methods are not feasible if the state and action spaces are large. Recently, online methods are actively researched such as Monte Carlo Tree Search(MCTS)(Silver and Veness 2010). MCTS gets popularity once it has been applied to go game successfully(Coulom 2007; Gelly et al. 2006). MCTS looks ahead with possible actions and observations by simulations. The possibility of improving the speed of simulations in MCTS using techniques of rare event simulation will be studied.

Human Modeling

Since the pilot of the UAV is human, a bounded rational human model is required to have realistic simulation results. There are a few methods to model human behavior. These methods define how to select actions.

ϵ-greedy method An action of maximum expected utility is selected with $1 - \epsilon$ probability. Other actions are selected randomly with ϵ probability.

Roulette method An action is selected probabilistically based on the ratio of utilities over all actions as shown in Eq 4.

$$P(a) = \frac{U(a)}{\sum_{\forall a} U(a)} \tag{4}$$

Boltzmann distribution method An action is selected probabilistically based on the Boltzmann distribution shown in Eq 5.

$$P(a) = \frac{\exp(U(a)/T)}{\sum_{\forall a} \exp(U(a)/T)} \quad (5)$$

T is a temperature parameter that controls randomness.

Multi-Players Cooperative Game

Finally, the UAV pilot and manned aircraft pilot are modeled as human. They are collaborating to achieve common goals of safety and surveillance. They have uncertainty about both the state of the environment and the choices of the other agent. Each agent needs to reason about the other and execute its own policy. Decentralized-POMDPs (Seuken and Zilberstein 2005; Oliehoek 2012; Goldman and Zilberstein 2004) provide a framework to model this kind of problems. This type of problems is usually extremely difficult to solve but there are several approximate solution methods.

Level-k Model The level-k model (Camerer 2003) provides a reasoning model about other human players. When building a decision making system that interacts with humans, computing the Nash equilibrium is not always helpful. Humans often do not play a Nash equilibrium strategy due to cognitive limitations. Level-k models assume that humans are erroneous and limited in the number of steps of strategic look-ahead, and it works well in practice. In the model, a level-k agent assumes the other agents adopt level-1 strategies and select actions according to the logit distribution.

Joint Equilibrium Search for Policies (JESP) JESP (Nair et al. 2003) finds a Nash equilibria in the cooperative game represented as the Dec-POMDP. It utilizes alternating best response. Policies of all but one are fixed and the remaining agent computes a best response to the fixed policies. This process is performed for every agent and repeated until no agents change their policies.

Max-n Monte Carlo Search Max-n Monte Carlo Search(Samothrakis, Robles, and Lucas 2011) is the method applying Monte Carlo Tree Search to max-n game tree. This method has been applied to Pac-Man game successfully. Max-n game tree is an n-player game tree with nodes represented as a tuple of utilities of all agents. Agents choose actions that maximize their own utility. MCTS is applied to prune the game tree so that the optimal path in the tree can be computed efficiently.

Preliminary Study

This section introduces a wildfire surveillance scenario. The scenario has been iterated multiple times with feedback from firefighters. A stochastic wildfire propagation model has been used. This preliminary study analyzes how uncertainties and communication loss influence the decision of the UAV in the scenario.

Scenario and Modeling

Figure 1 shows the scenario visually. The wildfire area is modeled as a 11×11 grid world. There is a wildfire in the center of the grid and the fire propagates stochastically.

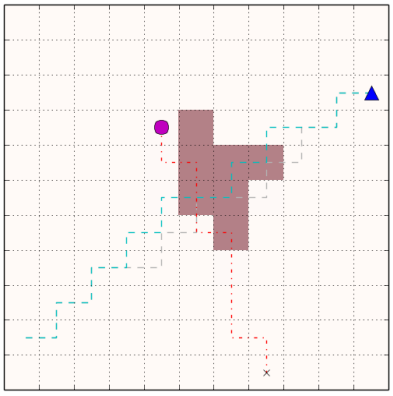


Figure 1: Scenario

There is one UAV (circle) and one manned aircraft (triangle). Both the unmanned aircraft and manned aircraft monitor the area. Communication between the UAV and UAV pilot is lost due to either hardware failure or radio inference.

The manned aircraft follows a planned path with a certain noise. The gray dotted line is the planned path and the blue dotted line is the actual path. The UAV can choose one of three actions when the communication is lost and the decision does not change until the end of simulations. It is a single-shot decision problem.

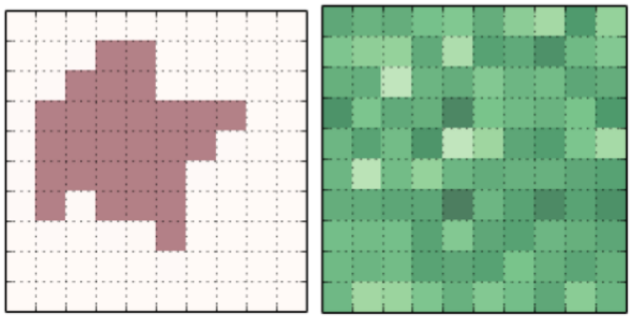
- back to base
- emergency landing
- stay in place

In the figure, the x mark is the location of the base and the red dotted line is the path. For the emergency landing action, the unmanned aircraft lands immediately at the current location. It can crash when it lands on fire or by a certain chance even when it does not land on fire. The last option is that the unmanned aircraft stays in place until the manned aircraft leaves the area.

Negative rewards are given if the unmanned aircraft crashes or comes too close to the manned aircraft. The total utility is the sum of rewards accrued during a simulation.

Extension of Scenario

The first scenario involves determining the best action when communication is lost. The scenario has been extended to include simulations after the communication returns. Positive reward is given when the UAV monitors the area after the communication returns. In this extended scenario, the manned aircraft chases the rim of fire and unmanned aircraft follows a lawn mower pattern for surveillance after the communication returns. A new action for the UAV is introduced. The UAV can lower its altitude until the manned aircraft leaves the area.



(a) burning matrix, B (b) fuel matrix, F

Figure 2: Wildfire Model Variables

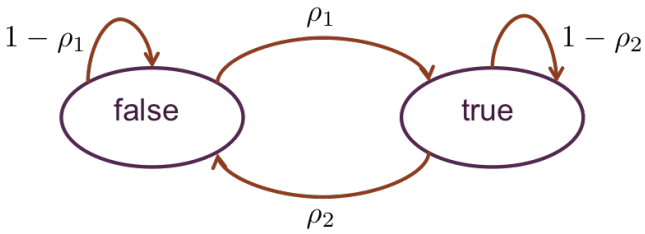


Figure 3: Transition of burning matrix, B

Wildfire Propagation Model

A simple stochastic wildfire model (Bertsimas et al. 2014) is chosen for simulations. There two model variables B and F. For each location x in the grid, $B(x)$ and $F(x)$ indicate whether the cell is burning or not and how much fuel is remaining in the cell. The figure 2 shows an example of B and F. Fire propagates probabilistically based on $p(x, y)$, which is the probability that a fire in cell y ignites a fire in cell x. Transitions of B and F are described in Figure 3 and Eq 6 7 8.

$$\rho_1 = \begin{cases} 1 - \prod_y (1 - P(x,y)B_t(y)) & \text{if } F_t(x) > 0 \\ 0 & \text{o.w.} \end{cases} \quad (6)$$

$$\rho_2 = \begin{cases} 1 & \text{if } F_t(x) = 0 \\ 0 & \text{o.w.} \end{cases} \quad (7)$$

$$F_{t+1}(x) = \begin{cases} F_t(x) & \text{if } B(x) = \text{false or } F_t(x) = 0 \\ F_t(x) - 1 & \text{o.w.} \end{cases} \quad (8)$$

Simulation

Expected utilities are calculated at every location for every action of the unmanned aircraft. For each pair of location and action, multiple simulations are performed until the average of expected utility converges. Figure 4 shows a simulation of a UAV initially located at $(4, 5)$ and it chooses back to base action during the communication loss.

Figure 4: Simulation

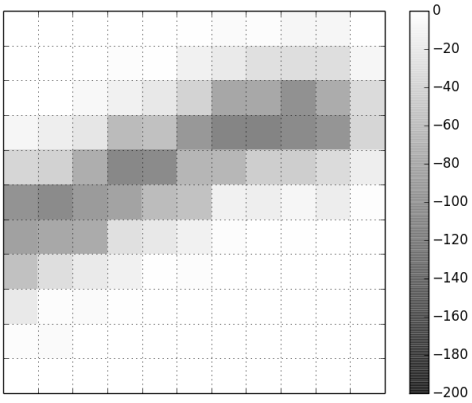


Figure 5: Expected Utility Map

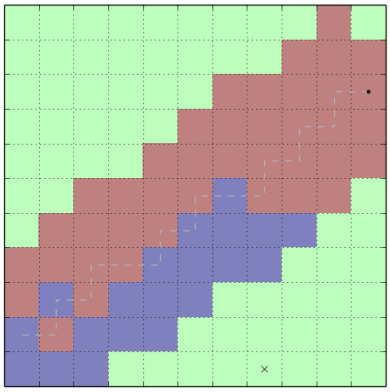
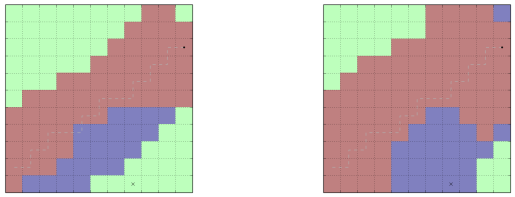


Figure 6: Policy Map



(a) small uncertainty (b) large uncertainty

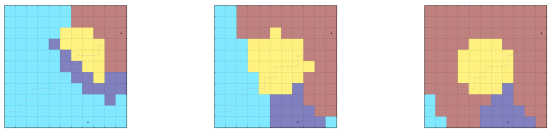
Figure 7: Policy map with manned aircraft path uncertainties

Results

Initial scenario Figure 5 shows a map of expected utilities for back to base. For example, an unmanned aircraft located at $(4, 8)$ gets a large negative utility if it chooses the back to base action when the communication is lost. This large negative utility is because the unmanned aircraft encounters the manned aircraft on the way back to the base. There is a different utility map for each action. For a given location of unmanned aircraft, the best action is the one that has a maximum utility among utilities of all actions at the location.

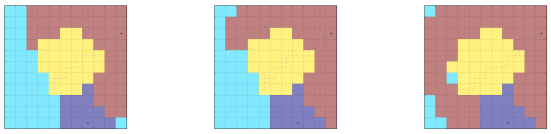
Figure 6 shows a policy map. For each location, the map shows which action is best. Blue, green, and red colors represent back to base, stay in place and emergency landing actions respectively. As shown in the figure, emergency landing action is the best for locations in the path of manned aircraft. Otherwise, unmanned aircraft stays in place or is back to base. Figure 7 shows policy maps as increasing the uncertainty of manned aircraft path. Red area gets larger because manned aircraft deviates more from the planned path as the uncertainty increases.

Extended scenario Figure 8 and Figure 9 show how the policy map changes as the duration of communication loss varies and how the uncertainty of communication loss duration influences the policy. Blue, cyan, red, and yellow colors represent back to base, stay in place, emergency landing and lower altitude actions, respectively. As shown in the figures, the optimal action is greatly impacted by communication loss duration and its uncertainty.



(a) 5 time steps (b) 10 time steps (c) 15 time steps

Figure 8: Policy map with various durations of communication loss



(a) std. 1. (b) std. 2. (c) std. 5.

Figure 9: Policy map with various standard deviations of communication loss duration

In the figures, the impact of communication loss has been studied without considering the surveillance reward after the communication returns. Figure 10 shows how the surveillance reward impacts the optimal action. Ten time steps is chosen for the communication loss duration. Thus, Figure 10 (a) is the same as Figure 8 (b) without surveillance reward. As the surveillance reward increases, lower altitude or stay in place action is preferred over emergency landing or the back to base action. This is because unmanned aircraft gets surveillance reward after the communication comes back, whereas it does not with emergency landing or back to base action.

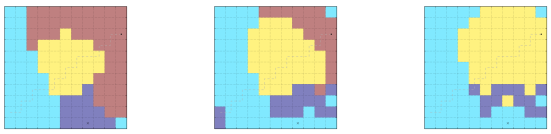
Future Work

Fast-Time Simulation

Since crashes and failed landings are rare, the current simulation framework requires a lot of time to obtain accurate simulation results. Techniques such as importance sampling or cross-entropy method will be investigated to make the simulation faster.

Sequential Decision of UAV Pilot

The UAV pilot makes a decision once right after the communication returns in the previous scenario. This scenario can



(a) reward 0. (b) reward 0.5 (c) reward 1.

Figure 10: Policy map with various surveillance rewards

be extended to model multiple communication losses with stochastic start and end times. The UAV pilot needs to make decisions to make unmanned aircraft avoid a collision with manned aircraft base on noisy information about the location of manned aircraft. Future work will explore MDP and POMDP formulations.

Model UAV Pilot as Human

In the preliminary study, the UAV pilot is assumed to command the unmanned aircraft to follow a lawn mower pattern for surveillance. In reality, the UAV pilot has multiple choices about the control of the unmanned aircraft. Moreover, since the UAV pilot is human, a bounded rational model for human is required.

Two-Players Game

The UAV pilot is human, although the pilot is assumed to be an intelligent robot in the preliminary study. The scenario is to be a cooperative two-player game between manned aircraft pilot and unmanned aircraft pilot. A Dec-POMDP or two-player cooperative game will be studied to formulate the problem and some techniques such as level-k, JESP, or max-n Monte Carlo tree search will be applied to this problem.

References

Bertsimas, D.; Griffith, J. D.; Gupta, V.; Kochenderfer, M. J.; Mišić, V. V.; and Moss, R. 2014. A comparison of monte carlo tree search and mathematical optimization for large scale dynamic resource allocation. *arXiv preprint arXiv:1405.5498*.

Camerer, C. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Coulom, R. 2007. Efficient selectivity and backup operators in monte-carlo tree search. In *Computers and Games*. Springer. 72–83.

De Boer, P.-T.; Kroese, D. P.; Mannor, S.; and Rubinstein, R. Y. 2005. A tutorial on the cross-entropy method. *Annals of Operations Research* 134(1):19–67.

Gelly, S.; Wang, Y.; Munos, R.; and Teytaud, O. 2006. Modification of uct with patterns in monte-carlo go.

Goldman, C. V., and Zilberstein, S. 2004. Decentralized control of cooperative systems: Categorization and complexity analysis. *Journal of Artificial Intelligence Research (JAIR)* 22:143–174.

Hauskrecht, M. 2000. Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research* 33–94.

Kim, Y. J., and Kochenderfer, M. 2015. Improving aircraft collision risk estimation using the cross-entropy method. In *AIAA SciTech*. American Institute of Aeronautics and Astronautics.

Kochenderfer, M.; Espindle, L.; Kuchar, J.; and Griffith, J. D. 2008. Correlated encounter model for cooperative aircraft in the national airspace system version 1.0. *Project Report ATC-344, Lincoln Laboratory*.

Kochenderfer, M. J.; M. Edwards, M. W.; Espindle, L. P.; Kuchar, J. K.; and Griffith, J. D. 2010. Airspace encounter models for estimating collision risk. *Journal of Guidance, Control, and Dynamics* 33(2):487–499.

Kochenderfer, M. J. 2015. *Decision Making Under Uncertainty: Theory and Application*. MIT Press.

Kurniawati, H.; Hsu, D.; and Lee, W. S. 2008. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems*, volume 2008. Zurich, Switzerland.

Nair, R.; Tambe, M.; Yokoo, M.; Pynadath, D.; and Marsella, S. 2003. Taming decentralized pomdps: Towards efficient policy computation for multiagent settings. In *IJCAI*, 705–711.

Oliehoek, F. A. 2012. Decentralized pomdps. In *Reinforcement Learning*. Springer. 471–503.

Pineau, J.; Gordon, G.; and Thrun, S. 2006. Anytime point-based approximations for large pomdps. *Journal of Artificial Intelligence Research* 335–380.

Rubinstein, R. Y., and Kroese, D. P. 2004. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer.

Samothrakis, S.; Robles, D.; and Lucas, S. 2011. Fast approximate max-n monte carlo tree search for ms pac-man. *Computational Intelligence and AI in Games, IEEE Transactions on* 3(2):142–154.

Seuken, S., and Zilberstein, S. 2005. Formal models and algorithms for decentralized control of multiple agents. *University of Masschusetts, Amherst, Computer Science Department, Tech. Rep*.

Shani, G.; Pineau, J.; and Kaplow, R. 2013. A survey of point-based pomdp solvers. *Autonomous Agents and Multi-Agent Systems* 27(1):1–51.

Silver, D., and Veness, J. 2010. Monte-carlo planning in large pomdps. In *Advances in Neural Information Processing Systems*, 2164–2172.

Smith, T., and Simmons, R. 2004. Heuristic search value iteration for pomdps. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 520–527. AUAI Press.

Whitlock, C. 2014. When drones fall from the sky. *The Washington Post*.