



HUMAN SWARM MODEL DESCRIPTION

Skanda Vaidyanath
Institute of Creative Technologies
05/30/2019

KEY POINTS

- The aim is to have a simplified state space and an action space to make sure that we learn how to interrupt the operator at the right times.
- For this reason, we will not allow the spokesperson i.e. the agent to make certain actions such as Search, Overview and Move. They blow up the action space (and the state space) as we need to specify the drone number and the locations.
- The operator has access to the map and the entire set of actions. He/she can call any action that he/she wants.
- We do not consider the actions of Confirm Observation and Remove Obstruction.
- Guide play simply makes a drone take a willing civilian group to the safe zone. We do not consider cases of humans going off course or obstacles in the way.

KEY POINTS

- We discretize the state variables to make sure that we do not deal with a continuous state space and to reduce the number of states.
- All the terms used and plays called are exactly in line with the MxR demo and the Confluence page. The grid is 32×32 as described by MxR.
- When the SP interrupts the operator, it gives the operator information about the state space and the operator can use his/her discretion to do what he/she wants to. It does not tell the operator what to do.
- As the SP is controlling the drones, the operator is also simultaneously making plays. The environment is also non-stationary. So the state could change even when the SP does nothing. To formalize this as an MDP, we add a NOP action.
- While the SP is negotiating through a drone, it cannot carry out any other simultaneous negotiation but all other plays are allowed.

STATE VARIABLES

- Locations of fire: 4 bit-vector. Divide the map into 4 equal 16×16 quadrants. If at least one sector of the quadrant is on fire, the bit for that quadrant is set to 1. The remaining are 0. (16 possibilities)
- Drone d status variables: ($1 \leq d \leq 10$)
 1. 0 → Offline(Dead)/ Idle/ Moving/ Tracking Fire(in Overview)/ Searching
 2. 1 → Warning
 3. 2 → SP negotiating
 4. 3 → Operator negotiating
 5. 4 → Monitoring
 6. 5 → Group convinced. Guide call available.
 7. 6 → Life Found (the drone is also by default in the search state, but this takes precedence.)
 8. 7 → Guiding(8^{10} possibilities)

STATE VARIABLES

- Drone d engagement variables ($1 \leq d \leq 10$):
 1. Number of warnings issued in current engagement (max value 2 i.e. counter stops at 2)
 2. SP negotiations issued? (0 for no, 1 for yes.)
(6^{10} possibilities)

Current engagement means the current civilian group that the drone is engaging with.

STATE SPACE ANALYSIS

- The number of states is nearly 10^{18} .
- The bottleneck is the number of drone specific variables.
- It is difficult to reduce the number of variables any further but reducing the number of drones from 10 to 5, gives us a state space in the order of 10^9 . If we have only 3 drones, it is in the order of 10^6 .
- Should we make only a subset of drones in the control of the SP?

ACTIONS (PLAYS)

- Guide(Drone d): A drone that is ready to guide (status 3) starts guiding the civilian group to the safe zone. If played correctly, gets a reward of +10. If the drone is not ready to guide, this incurs a reward of -10 and nothing happens.
- Engage(Drone d): A drone that has spotted life (status 4), can engage with that civilian group. Drone then goes into a monitoring state. It gets a reward of +10 if done successfully. If this action is called when the drone is not in Life Found status, this leads to a reward of -10 and nothing happens.
- Warn(Drone d): Drone d issues a warning and this can be called only if the drone is in the monitoring state (status 2). Incurs a reward of -3 if done successfully. If called when drone is not in monitoring state, gets a reward of -10.

ACTIONS (PLAYS)

- SP Negotiation(Drone d): SP negotiates through this drone and this can be called only if the drone is in the monitoring state (status 2). Incurs a reward of -10 if done successfully. If called when drone is not in monitoring state, gets a reward of -20. If called when any other drone is in SP negotiating status, -20 and nothing happens.
- Interrupt Operator: The SP interrupts the operator and incurs a reward of -25. This can be done at any time. The operator looks at the map and decides what needs to be done. If called when any other drone is in operator negotiating status, -40 but the operator is interrupted again. The operator decides what to do in this scenario and the state may change accordingly.
- NOP: To account for state changes that happen when the SP doesn't carry out any action.

ACTION SPACE ANALYSIS

- There are a total of 42 actions since the first four actions can be carried out by each of the 10 drones separately. If we reduce the number of drones to 5, then we have 22 actions. For 3 drones, only 14 actions.
- If we add actions for search and move as well, we will likely need to include location information for each drone in the state variables which would blow up the state space.
- Additionally for search and move, we would have to specify the drone number and the location to search/ move to. This would lead to an extremely large number of actions.



REWARDS

- Civilian Group to safe zone(saved): +5000
- Civilian Group dead: -500,000

HOW DO WE HANDLE STATE CHANGES?

- Here is an example scenario.
- Assume we have a state where a drone D is in the monitoring state (status 2). We call the warning play, taking us into a new state where now, the drone is in the interacting state (status 1) and the counter for the number of warnings issued in the current engagement increases if possible. We incur a reward of -3 for the action as well. This is the transition from $s(t)$ to $s(t+1)$.
- Now when the warning has been completed (there may be a time taken to issue the warning), the drone D goes back to the monitoring state. There is a state change again but since there was no explicit action from the SP, we label the action as NOP. We do not get any reward. We transition from $s(t+1)$ to $s(t+2)$.

HOW DO WE HANDLE STATE CHANGES?

- So we go from $s(t)$ (M state) $\rightarrow s(t+1)$ (W state) $\rightarrow s(t+2)$ (M state) where M stands for drone monitoring and W stands for drone warning. These states are with respect to the specific drone D in engagement.
- Note that we may stay in the W state for a while depending on how long the warning takes. The SP can also make other plays in the mean time, changing the state of the system. With respect to the engaging drone D, it is still in the W state.
- It is also possible that the operator might make some changes that affects our state and we will follow the transitions with NOP actions. Still, with respect to drone D, we are still in the W state. But the actual state of the system may change. (Unless the operator reassigns drone D. In that case the warning is interrupted and the drone D goes away to an entirely new state)
- Eventually a NOP will take us to $s(t+2)$ and then we can play another action. Note that none of the SP's actions are valid i.e. they don't do anything in the W state (for drone D).