Mixed Initiative Balance of Human-Swarm Teaming in Surveillance Via Reinforcement Learning

Tristan Scheiner and Luke Ferderer

Introduction

- In certain types of missions, a UAV swarm is shown to have higher efficiency than a single UAV agent including package delivery, emergency rescue, and field surveillance
- The quantities of information exchange of UAV to UAV and UAV to human should not be the same (all information about the states of individual UAVs should not all be sent to the human operator)
- The trust of a human to the swarm is based on several factors including workload, situation awareness, and to the trust to swarm teammates

Contributions

- Computational mode of human→swarm trust
- Use Q-Learning to optimize for human→swarm trust
- Design of a robust surveillance scheme

Related Works

- Other works have tried to quantify “trust” between humans and swarms
  - The more clear a swarm’s actions are to humans, the more they tend to trust the swarms
- Multiple models of robot control and behavior have been investigated to improve human trust in the robotics system
- Swarm state needs to be abstracted for humans to understand

Communication Aware UAV Swarm in Persistent Surveillance

- Agents in the swarm communicate with neighboring agents and the human operator through the use of simulated multiple input and multiple output antenna
- Physical model of the UAV swarm communication neglects co-channel interference
- Assume that co-channel signal and environmental noise is suppressed by signal processing algorithms
- Environmental noise may result in lost data packages from being transmitted to their neighbors

Computational Model of HSI

- Multiple metrics are calculated to define models of trust
- Trust models are defined for the reinforcement learning (RL) model:
  - Human to Swarm trust ($T_{h \rightarrow sm}$)
  - Swarm to Human trust ($T_{sm \rightarrow h}$)
    - The total $T_{sm \rightarrow h}$ is the average trust of each swarm agent in the human operator
- As required by the RL algorithm, both trust metrics are calculated at each time step
- Finally, a swarm utilization rate is calculated for evaluation

Reinforcement Learning Based HSI Mixed Initiative Balance

- Different methods to influence the trust between the human and the swarm include:
  - Increasing information feed to human workers
  - Increasing or reducing thrust to match mission disturbance
  - Adopting a robust communication scheme to provide extra robustness for the team
- Intend to map human to swarm trust to UAV control with the use of reinforcement learning
- Agent chooses action based on current human to swarm trust

What is Q-Learning

- Model-free, off-policy method to maximize a reward based on actions that an agent takes in the environment
  - Model-free = Does not use the state transition model when training
  - Off-policy = Learning does not happen with the same policy that was used during exploration
- Q-values (predicted expected return of a state) are represented in a 2D table where one axis are states and the other axis are actions
  - Only feasible to use Q-learning on a low discrete action and state space environment

REWARD TABLE OF TRUST AND UAV CONTROL

State	Drop Thrust	Increase Thrust	Keep Thrust
(FT,3)	-100	1	50
(FT,5)	20	-100	5
(FT,7)	20	-200	1
(LT,5)	-200	50	1
(LT,10)	1	20	5
(LT,20)	1	5	20
(NT,10)	-200	50	-100
(NT,20)	-100	20	1
(NT,35)	20	1	10

Simulation Results and Discussion

- Simulation of 6 checkpoint positions the swarm should reach
- Simulated a baseline without Q-learning
- Models used to calculate robot thrust based on trust

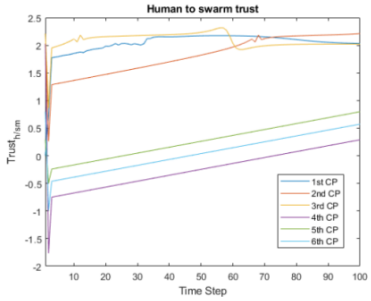

Fig. 4. Human to Swarm Trust (No HSI Adjustment)

Without Q-learning, T_h→sm is highly varied between checkpoints

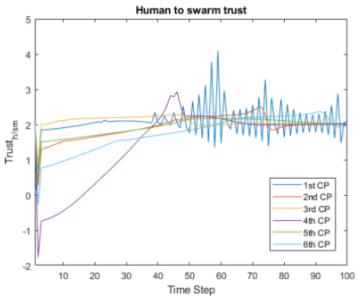

Fig. 9. Human to Swarm Trust based on HSI

Conclusion

- Paper proposed a dynamical mixed initiative model for human to swarm trust interaction
- Both human to swarm and swarm to human trust models are built
- Q-learning is used to final the optimal thrust based on trust values
- Proposed to use online DRL methods in the future

Hard to Read (We promise context doesn't help)

- "Sean et al. investigate the mechanism of rebuild the human to swarm trust after the robot teammate has made mistakes."
- "The results are better results of estimation compared to dynamic Bayesian networks"
- "which is the value of human makes the shortest and longest move in one time step"

Hard to Read Cont. (We promise context doesn't help)

- "The exponential component describes the human's memory to his utilization history of mission execution. The ratio of moving distance in a one-time step shows humans' current utilization rate"
- "In the preset human swarm cooperation scenario, results show that this method has less affection for humans to predict the future state of the swarm compared to the complete state of the swarm shown to the human operator."
- "Swarm to human trust is computed with the UAV agent to human trust which is dependent on the waypoint tracking performance of the human operator."
- "... how the strategies allocation on the high level of decision making could be effectively described based on the human swarm interaction."
- "Human mission performance is defined when worker's surveillance route follows a linear function."

Hard to Read Cont. 2 (We promise context doesn't help)

- "Depends on the mission, human operators perform multiple types of duties to ensure the performance of UAV teammate, and robustness of this human-in-loop system."
- "Third, CP has the most anomaly value caused by the sudden decreasing moving distance from 2nd to 3rd CP."
- "Inspired by the human-to-human cooperation, HST trust model, originating from human-machine teaming, trust model with HSI is usually considered to depict the human-machine relation in a mission."

Houbing Song, the Mystery Man

- Ph.D in 2012
- Received his Ph.D in 2012
- 575 publications according to Research Gate (~52.27 a year)
- An IEEE Fellow and ACM Distinguished Scientist
- Head of the University of Maryland, Baltimore County's SONG lab
  - Currently lists 7 Ph.D students and 23 alumni
  - Assuming his name is on every paper an active Ph.D student makes, it averages 1 paper every 2 months per person
- Won best paper awards in conferences in the following subjects:
  - Cyber, Physical and Social Computing
  - Industrial Internet
  - Cloud and Big Data Computing
  - Wireless Algorithms, Systems, and Applications