Towards Responsible Human-Agent Collectives

Sarvapali D. (Gopal) Ramchurn
Professor of Artificial Intelligence
Director, Centre for Machine Intelligence

Invited Talk, Thessaloniki, October 2018
Background

• Professor of AI (2018-) at the University of Southampton
• Director, Centre for Machine Intelligence (2017-)
  • 130+ Researchers
  • Top Group for Impact in the UK
  • 15+ Projects in AI and Machine Learning
• Academic
  • PhD in Multi-Agent Trust and Negotiation (Southampton) – 2004

• Advisory roles:
  • UK Cabinet Office (OI team) (2018-)
  • Tech Startups: Engagetech and UTU Kenya
  • Chief Scientist at North Star Solar (2017-)
• Awards:
  • Best Paper/Nominations AAMAS 2010/11,13,15, IJCAI-JAIR
  • AXA Award for Responsible AI (2018)
Projects

• Past Research Projects:
  • ALADDIN (EPSRC – BAE 2005-2010)
  • IDEAS: Intelligent Decentralised Energy-aware Systems (SECURE LTD 2009-2013)
  • ORCHID: Foundations of Human-Agent Collectives (EPSRC Programme 2011-2016)
  • SEACORES: Fault-Diagnosis on Ships (I-UK 2015-2016)
  • CharIoT: Energy monitoring kit for energy advisors (EPSRC 2015-2017)
  • Human-UAV Teaming (2015-2017)

• Current Projects
  • AXA Responsible AI (2018-2021) £200K
  • Smart Cities and Wearable Tech (EPSRC 2017-2021) £1.2M
  • Autonomous IoT (EPSRC 2016-2019) £800K
  • GCRF BRECCIA (EPSRC 2018-2022) £2.5M
Goals of this Talk

• Explain what human-agent collectives (HACs) are
• Detail some examples of HACs and Applications
• Present some new research directions for Responsible AI/HACs
An Era of Information Ubiquity

Connected
Diverse Sources
Measure everything
Mashed up

During 2008, the number of things connected to the internet exceeded the number of people on earth.

By 2020 there will be 50 billion.
More helpful computers for a new way of life

(Jennings et al., 2014)

Proactive

Mixed-reality

Machines as collaborators

Human-Agent Collectives (HACs)
Flexible Autonomy

neither agents, nor humans are always in charge

• Humans act with varying degrees of computer support.
• Agents can act autonomously, other times guided by much closer human involvement.
• Vary depending on context.
Agile Teaming
continually establish and manage collaborative relationships

Groups of agents and humans:

• **Come together** when needed to achieve goals no individual can achieve in isolation

• **Disband** once cooperative action has been successful.
Incentive Engineering
motivate by incentive, rather than diktat

• Design rewards so actions that are encouraged generate desirable outcomes.
• Account for human perception and motivations
Accountable Information Infrastructure

track information veracity and provenance

- heterogeneous data that has varying degrees of reliability and accuracy.
- Allow veracity and accuracy of information to be confirmed and audited, while maintaining privacy and ethics standards.
HACs in Smart Energy Systems

- Smart Heating Control
- Personalised Recommendations & Advice Giving
- Electric Vehicle Charging
HACs in Citizen Science

Classifying Galaxies

Hunting for Endangered Species
HACs in Disaster Response
12/01/2010, Port au Prince, Haiti 2010

HAITI Stats
- 230,000--316,000: estimates of the death toll vary.
- 300,000: number of injured
- 1.5 million: people initially displaced
- 85,432: displaced people remain in 123 sites as of September 2014

Response in Dollars: $13.34 billion
A Disaster Response System based on HACs

Crowdsourcing and UAV Mission Planning for Situational Awareness

- Crowdsourced Data collection
- Classify & Summarise Crowd Reports
- Allocate UAVs to targets
- Send out UAVs

Rescue Mission planning

- Collect Data from UAVS
- Create Targets for First Responders
- Send out Responders
- Create an Allocation for First Responders

Provenance tracking of Decisions and Information
How do we **LOCATE** casualties and resources?

How to **ALLOCATE** resources?

How to **DEPLOY** rescue teams across a large area?

How do we **TRUST** the information gathered?
Information gathering and coordination loops in HAC-ER

Situational Awareness using Crowds and UAVs
- Crowdsourced Data collection
- Classify & Summarise Crowd Reports
- Allocate UAVs to targets
- Send out UAVs

Human-Agent Coordinated Task Allocation
- Collect Data from UAVS
- Create Targets for Bronze Responders
- Send out Responders
- Create an Allocation for Bronze Responders

Provenance of Decisions and Information
HAC-ER Modules

1. CrowdScanner
2. Mixed-Initiative Multi-UAV coordination
3. Mixed Initiative Task Planner
4. Provenance Tracker
CrowdScanner: making sense of crowd reports using human and machine intelligence

- **Interpretation**
  - Online (imperfect) Crowds + Machine Learning (BCC+ NLP)
  - Hire+Fire algorithm to recruit the best workers and get the best interpretation

- **Heatmap creation**
  - Gaussian Process to model disaster
  - Fold in trusted reports
  - Use classification output to generate intensity

- Generate targets for UAVs
Mixed Initiative Multi-UAV Coordination

• 1 pilot -> 3+ UAVs

• Heterogeneous UAVs running max-sum

• Flexible Autonomy
  • ‘Adjust’ max-sum plans
  • React to UAV drop-outs
  • Transfer of control between Silver, UAVs, and Bronze operators

• Validated on real UAVs

• Tested with 40 users in Lab

• UAVs Targets confirmed for Responders to be deployed
Human-Agent Collaboration for Task Allocation

- Human-Agent Silver team allocate tasks to Bronze responder team

- Agent uses Multi-agent Markov Decision Process
  - Computes best task for each responder, and best path for each task
  - Models environment (buildings and lakes are obstacles)

- Responders get instructions via mobile app

- Trialed in the AtomicOrchid Mixed Reality Game with 100+ users including emergency responders.
Supporting Human and Agent Decision Makers using Provenance

- Timely Decision Support
  - Live monitoring of provenance for changes
  - Ensures the whole system reacts to changes
- Post-hoc analysis

Example:
- During the operation, UAVs invalidate targets,
- Prov tracker immediately notifies Silver commanders at HQ
- Prov tracker identifies impacted rescue missions
HAC Interactional Arrangements

1. CrowdScanner
   - Multiple Humans as Sensors
   - Single Agent Learning

2. Mixed-Initiative Multi-UAV coordination
   - Multiple Humans Planning
   - Multiple Agents Planning
   - Embodied Agents

3. Mixed Initiative Task Planner
   - Multiple Humans Planning
   - Single Agent Planning

4. Provenance Tracker
   - Provenance Agent
   - Multiple Humans
   - Multiple Agents

CrowdScanner
Mixed-Initiative Multi-UAV coordination
Mixed Initiative Task Planner
Provenance Tracker
Human-UAV teaming in dynamic and uncertain environments
HAC Challenges

- **Organisational** challenges: processes and coordination mechanisms.
- **Interactional** challenges: interfaces and interaction modalities.
- **Accountability** challenges: human and agent as equal partners

Can we define a methodology to design HACs that are **Responsible**?
What is Responsible AI?

Controllable
Understandable
Trustworthy
Ethical
Reliable
Who makes optimal decisions?
Who is more precise?
Who makes the right decision?

Asimov’s 3 Rules Of Robotics

1. Robots must never harm human beings or through inaction allow a human being to come to harm.

2. Robots must always follow instructions from human beings unless they would cause them to violate rule one.

3. Robots must protect themselves unless it would cause them to violate the other two rules.
Model Checking for Responsible Swarms

- Define Actions and Consequences
- Ethical behaviour and moral dilemmas
  - Control Dilemma:
    - Let the user fly or not?
  - Save life and Infrastructure:
    - Crash the drone or damage property
    - Do not harm humans
- Decide under uncertainty and dynamism
Algorithms for Fair Load Shedding Problems

- Developing countries have an energy crisis
- Load shedding is essential
- Current load shedding techniques are not particular about fairness
- Can we use predictions of day-ahead consumption and supply to reduce unfairness?

<table>
<thead>
<tr>
<th></th>
<th>Grouper Algorithm</th>
<th>Consumption Sorter Algorithm</th>
<th>Random Selector Algorithm</th>
<th>Cost Sorter Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Minimize differences in disconnections</td>
<td>Minimize differences in disconnections &amp; supplied electricity</td>
<td>Minimize differences in disconnections &amp; supplied electricity</td>
<td>Minimize differences in comfort, disconnections &amp; supplied electricity</td>
</tr>
<tr>
<td>Description</td>
<td>Random grouping of households &amp; selection of group with least disconnections</td>
<td>Round-robin selection of households based on consumption</td>
<td>Round-robin selection of households in random order</td>
<td>Round-robin selection of households based on cost</td>
</tr>
</tbody>
</table>
Open Questions

• There are no metrics for Responsible AI
• Methodological:
  • Algorithms (e.g., avoid training bias, privacy preserving)
  • Interfaces (e.g., avoid automation bias)
  • Organisations (e.g., guarantee safety and ethical outcomes)
• Challenges:
  • Modelling humans
  • Evaluating interfaces and interactions
  • Incentives to change
  • Dealing with Ethics