EPSRC Engineering and Physical Sciences Research Council

BAE SYSTEMS

Bayesian learning for agent cooperation

M.Allen-Williams and N. R. Jennings
Intelligence, Agents, Multimedia Group
School of Electronics and Computer Science
University of Southampton

- An agent may need to **work with others** to achieve its individual goals
- If others are not cooperative, the agent must do its best to **fit in** with how others are behaving

Some tasks require cooperation

- Even if the others are cooperative, if the situation is complex, **computing optimal behaviour may be not be possible** in a timely fashion
- Especially as time or bandwidth constraints may **limit communication**

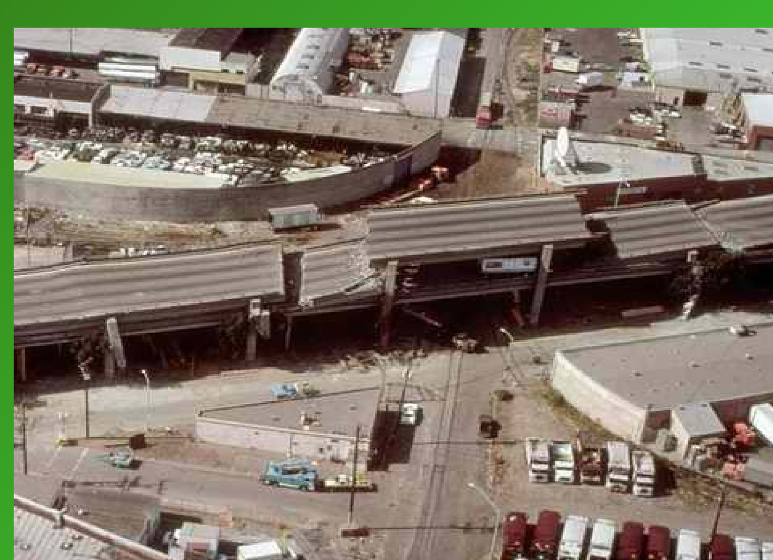
- Other agents will act according to:
  - their **strategy**, *and*
  - their **view** of the situation

Cooperating in uncertain environments means guessing what others are thinking

- If an agent is not aware of a disaster victim, it will not move to the rescue, even if the victim is close by and in critical condition
- We must model other agents' ***beliefs*** about the situation, as well as their strategies.

- After an earthquake, the state of the world is **uncertain**
- Aftershocks, fire and wind may **change the world** very rapidly
- Victims are buried under rubble and must be **found quickly**, **dug out** and treated

Example scenario



Earthquake in San Francisco

- Communications networks may be **down** or **congested**
- Rescuers from all areas and all directions must try and **coordinate** to discover and rescue the victims

- Agents can use **Bayes' rule** to update their beliefs about:
  - The state of the world s
  - The world dynamics θ
  - The strategies of the others σ_j
  - The observations and thus beliefs of the others o_j, b_j
- The agent maintains a **belief state** $P(M)$ over the model $M=(s, \theta, \{\sigma_j\}, \{o_j, b_j\})$

Bayes' rule

$P(Model \mid observations) \propto P(observations \mid Model)\, P(model)$

Guessing what other people are thinking is hard

- Recursive **Bellman equations** compute the **best response** action to a belief state, taking into account the effect on future states
- For large problems, the Bayesian update to M, and the best response, are **intractable**

Bellman equations

$$Q(b, a) = \Sigma_x P(x \mid b) \Sigma P(s' \mid x,a)[R(s') + \gamma V(b')]$$
$$V(b') = \max_a Q(b', a)$$

Terms

x : **unknowns** in model M
$R(s)$: immediate **reward** of world state s
$V(b)$: long-term **value** of belief state b
$Q(b, a)$: value **of action a** in belief state b

- *Internal states* of a **finite state machine (fsm)** determine the immediate action
- The subsequent **observations** determine the next internal state
- Maintaining beliefs over s, o_j and a set of fsms F_j is **tractable**

Approximate others, using *finite state machines*

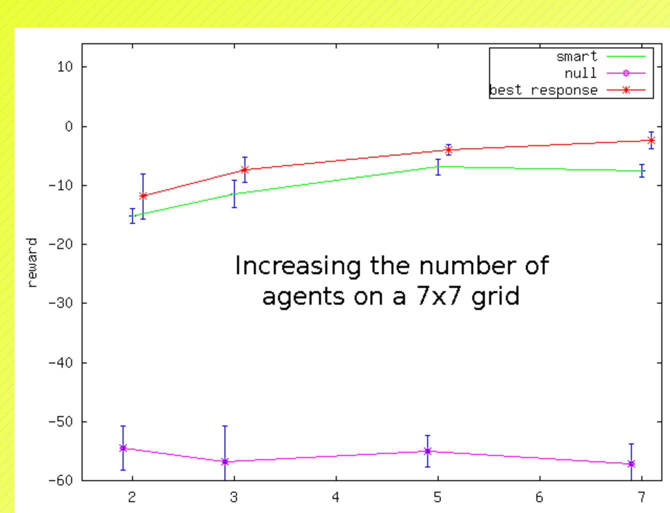
- Best response is still not tractable, because it iterates over **all possible belief-states**
- But this is unnecessary: look a few steps ahead, approximate the rest with a simple heuristic: **finite-horizon best response**

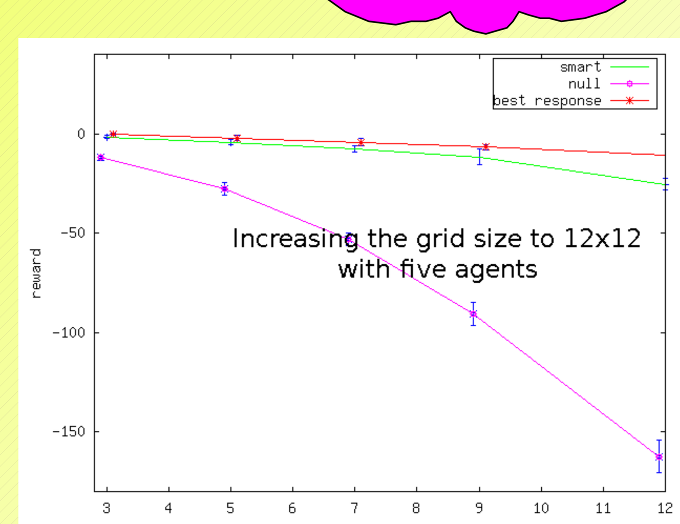
Finite-horizon best response (modified Bellman)

$$Q_{k+1}(b, a) = \Sigma_x P(x \mid b) \Sigma P(s' \mid x,a)[R(s') + \gamma V_k(b')]$$
$$V_k(b') = \max_a Q_k(b', a) \quad \text{and} \quad V_0(b') = heuristic(b')$$

- $n \times n$ **gridworld,** with k **ambulance agents**
- **Buried victims** continuously arriving across the grid
- **16^n** possible states: 4096 for $n = 3$, many millions for $n > 7$.
- Agents can **move** (r→, l←, u↑, d↓) or **dig**
- **5^k** *joint* actions: 125 for $k = 3$, 78,000 for $k = 7$

Success!





- Better than hand designed policy at capitalising on num. agents
- Better than hand designed policy as grid size increases

ALADDIN
autonomous learning agents for decentralised data and information networks

UNIVERSITY OF Southampton
School of Electronics and Computer Science