Towards concurrent Q-Learning With Local Rewards on Linked Multi-Component Robotic Systems

Borja Fernandez-Gauna, Jose Manuel Lopez-Guede, Manuel Graña

Computational Intelligence Group
University of the Basque Country (UPV/EHU)

IWINAC 2011, La Palma

Outline

Linked Multicomponent Robotic Systems

- Definition: group of robotic units physically-linked by a non-rigid element.
- Physical link introduces new non-linear dynamics and physical constraints in the system.
- Traditional control techniques are not appropriate

Multi-Agent Reinforcement Learning

- Reinforcement Learning (RL)
  - Set of algorithms that learn by exploring the state space S taking actions from set A
  - A reward function qualifies how good the observed state is $(R : S \rightarrow \mathbb{R})$
  - Goal: maximize the accumulated rewards over time
- Q-Learning
  - Estimates the rewards to be obtained after taking action a in state s by looking one step ahead:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \gamma * \max_{a'} \left\{ Q(s',a') - Q(s,a) \right\} \right)$$

Multi-Agent Reinforcement Learning

- Main RL drawback: exponential growth of the state-action space ($|S \times A|$)
- Multi-Agent Reinforcement Learning (MARL) makes it even worse: $|S \times A^n|$
- L-MCRS present an additional problem: physical constraints.
  - Some states force simulation to stop and start over
  - Examples:
    - physical-link stretched beyond its nominal length
    - collision between robotic units

Problem Statement

- A set of n linked robots (each of them represented as P_i) must carry the tip of a hose from a starting configuration to the goal
- Available actions: *Up, Down, Left, Right, Up-Left, Up-Right, Down-Left, Down-Right and None*
- Simple hose model: line segment
- Termination conditions:
  - A robot steps over the hose
  - Hose segments are stretched over nominal length
  - A robot gets out of the simulation world
  - Two robots colide
- Decentralized control and local rewards based on agents' selfish goal

Example

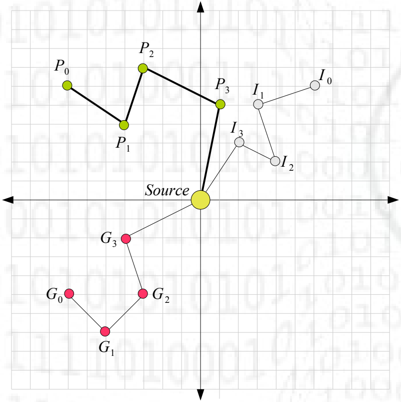

Figure: An example of the system: initial configuration (I_i), current position of the robots (P_i) and goal destination (G_i)

Multi-Agent Coordination

- For the agents to learn the best policy for each of the states, the straightforward approach uses omniscient agents
- State-action growth makes it unfeasible even in this simple environment
- Instead we use turns, so the state remains stationary during an agent's move
- Because of particularities of L-MCRS, we investigate the behavior of agents able to observe only a few state variables:
  - position of the agent and its neighbours
  - detection of an object in adjacent cells

Undesirable Termination Conditions

- Local reward function decomposition for each agent: a goal reward function $R^G : S \rightarrow \mathbb{R} \geq 0$ and several auxiliary functions $R_i^U : S \rightarrow \mathbb{R} \leq 0$
- R^G returns a positive reward whenever the goal is reached
- R_i^U return a negative reward when the i-th constraint is broken

State-Action Modular Veto approach

- Assuming not all R_i^U depend on all the state variables but a subset, the original problem can be decomposed in several concurrent modules
- One of them learns how to maximize R^G and the rest of modules learn state-action pairs leading to undesired terminations so as to veto them in the future
- The reduced state space makes considerably faster learning how to avoid them

State-Action Modular Veto approach

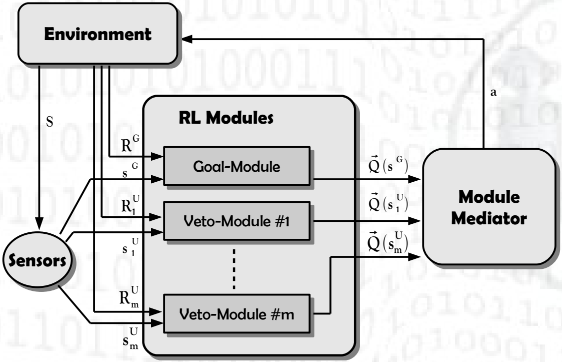

Figure: Scheme of the State-Action Modular Veto algorithm

Results

- Initial configurations were randomly generated
- One episode was simulated for each configuration with typical $\varepsilon - greedy$ exploration
- Percent of succesfull configurations was measured with a 500 episode window

Experiment A: No modular veto system

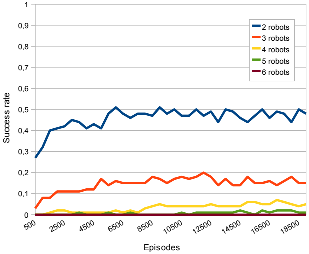

Figure: Results without the modular veto system

Experiment B: Modular veto system

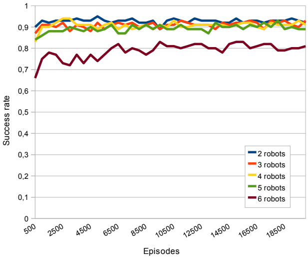

Figure: Results without the modular veto system

Thanks

Thank you very much for your attention.

- Contact:
  - Borja Fernández Gauna.
  - Computational Intelligence Group.
  - University of the Basque Country (UPV/EHU).
  - E-mail: **borja.fernandez@ehu.es**
  - Web page: **http://www.ehu.es/computationalintelligence**