Survey on Hierarchical and Modular Reinforcement Learning

Shun Zhang

September 26, 2014

Markov Decision Process

MDP:

- State: S.
- Action: A.
- Transition: $P : S \times A \times S \to \mathcal{R}$.
- Reward: $R : S \times A \times S \to \mathcal{R}$.

Abstraction on MDP

- Aggregate states: feature extraction.

Abstraction on MDP

- Aggregate states: feature extraction.
- Aggregate actions: **option**.

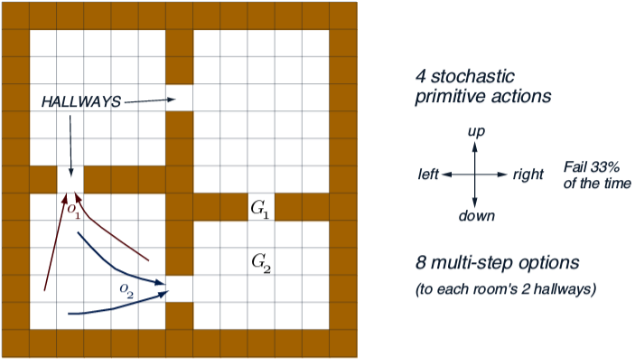
Abstraction on MDP

- Aggregate states: feature extraction.
- Aggregate actions: **option**.
- Decompose transition: factored MDP.

Abstraction on MDP

- Aggregate states: feature extraction.
- Aggregate actions: **option**.
- Decompose transition: factored MDP.
- Decompose value (abstract MDP): **HAM, hierarchical RL, modular RL**.

MDP with Option



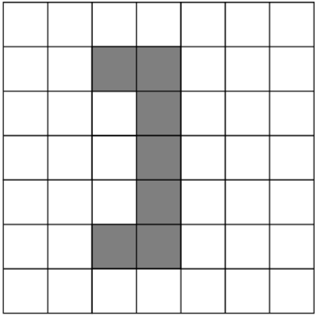
- Option: (start state, policy, termination condition).

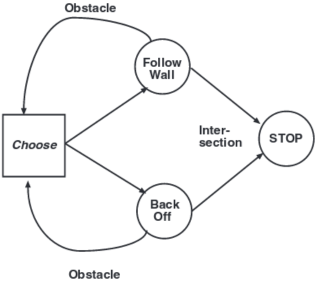
MDP with Option

MDP:

- State: S.
- Action: A, O.
- Transition: $P : S \times \{A, O\} \times S \to \mathcal{R}$.
- Reward: $R : S \times \{A, O\} \times S \to \mathcal{R}$.

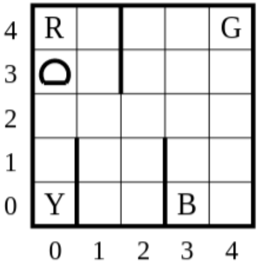
Hierarchies of Abstract Machines (HAM)

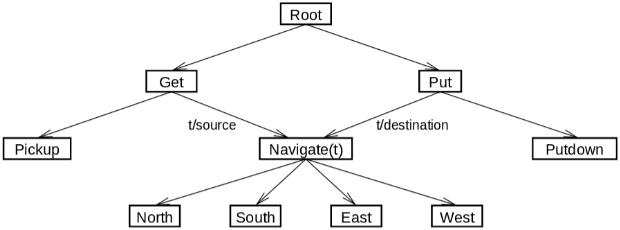




- State machine of MDPs.

Hierarchical RL





MDP:

- State: $\mathcal{S}$.
- Action: $\mathcal{A}$.
- Transition: $\mathcal{T}$.
- Reward: $\mathcal{R}$.

Modular RL

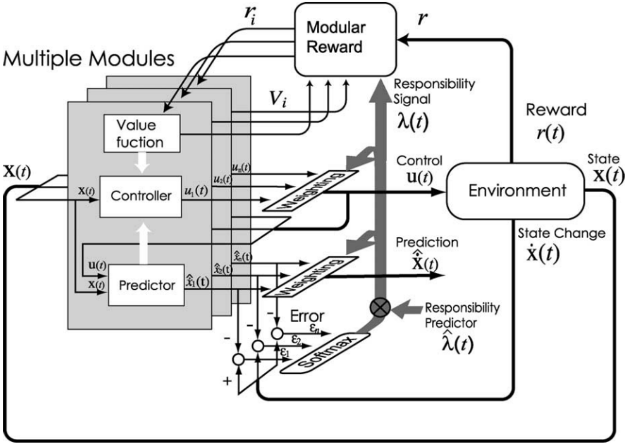


Fig. 1. MMRL.

MDP:

- State: $S_1 \times S_2 \cdots \times S_M$.
- Action: A.
- Transition: $P_1 \times P_2 \cdots \times P_M$.
- Reward: $R_1 \times R_2 \cdots \times R_M$.

Topics for Future Work

- Credit assignment.
- Learning task hierarchies.
- Dynamic abstraction.
- Integrating Deep Learning.