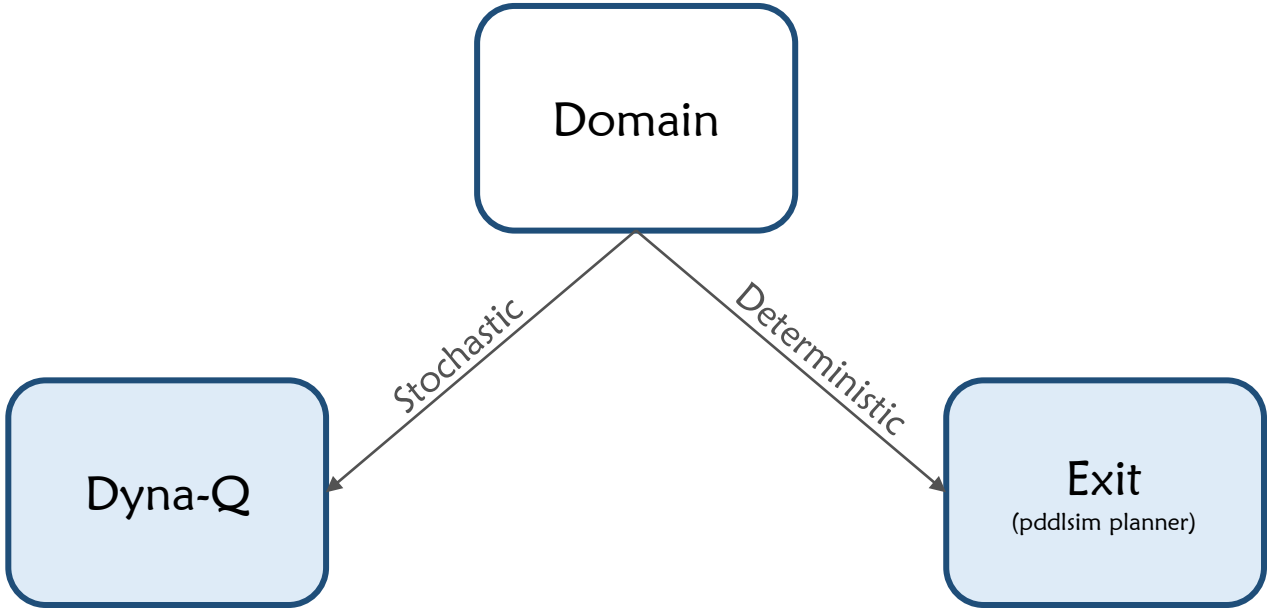
Introduction to Intelligent, Cognitive, and Knowledge-Based Systems

Final Project

Ayelet Tennenboim

Learning Phase



Dyna-Q

Dyna-Q Algorithm

Tabular Dyna-Q

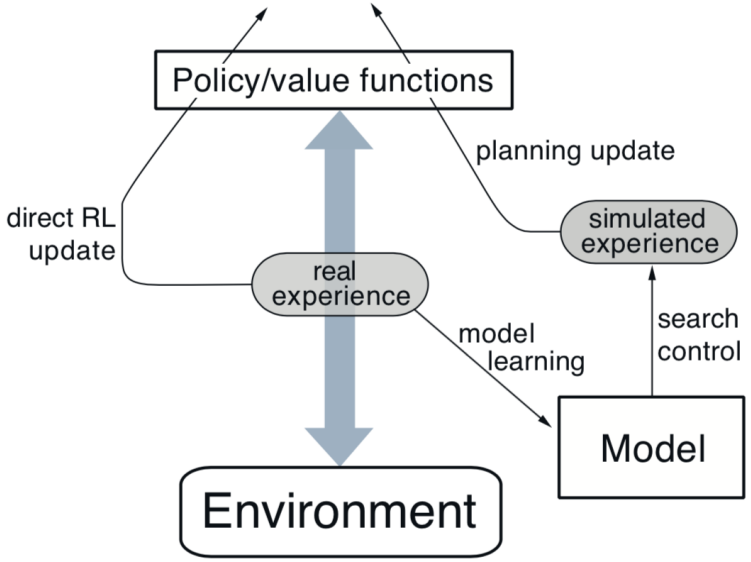
Initialize $Q(s,a)$ and $Model(s,a)$ for all $s \in \mathcal{S}$ and $a \in \mathcal{A}(s)$
Loop forever:
- (a) $S \leftarrow$ current (nonterminal) state
- (b) $A \leftarrow \varepsilon$-greedy(S,Q)
- (c) Take action A; observe resultant reward, R, and state, S'
- (d) $Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma \max_a Q(S',a) - Q(S,A)\big]$
- (e) $Model(S,A) \leftarrow R, S'$
- (f) Loop repeat n times:
  - $S \leftarrow$ random previously observed state
  - $A \leftarrow$ random action previously taken in S
  - $R, S' \leftarrow Model(S,A)$
  - $Q(S,A) \leftarrow Q(S,A) + \alpha\big[R + \gamma \max_a Q(S',a) - Q(S,A)\big]$

Q-Learning

Model Update

Planning Step

The general Dyna Architecture

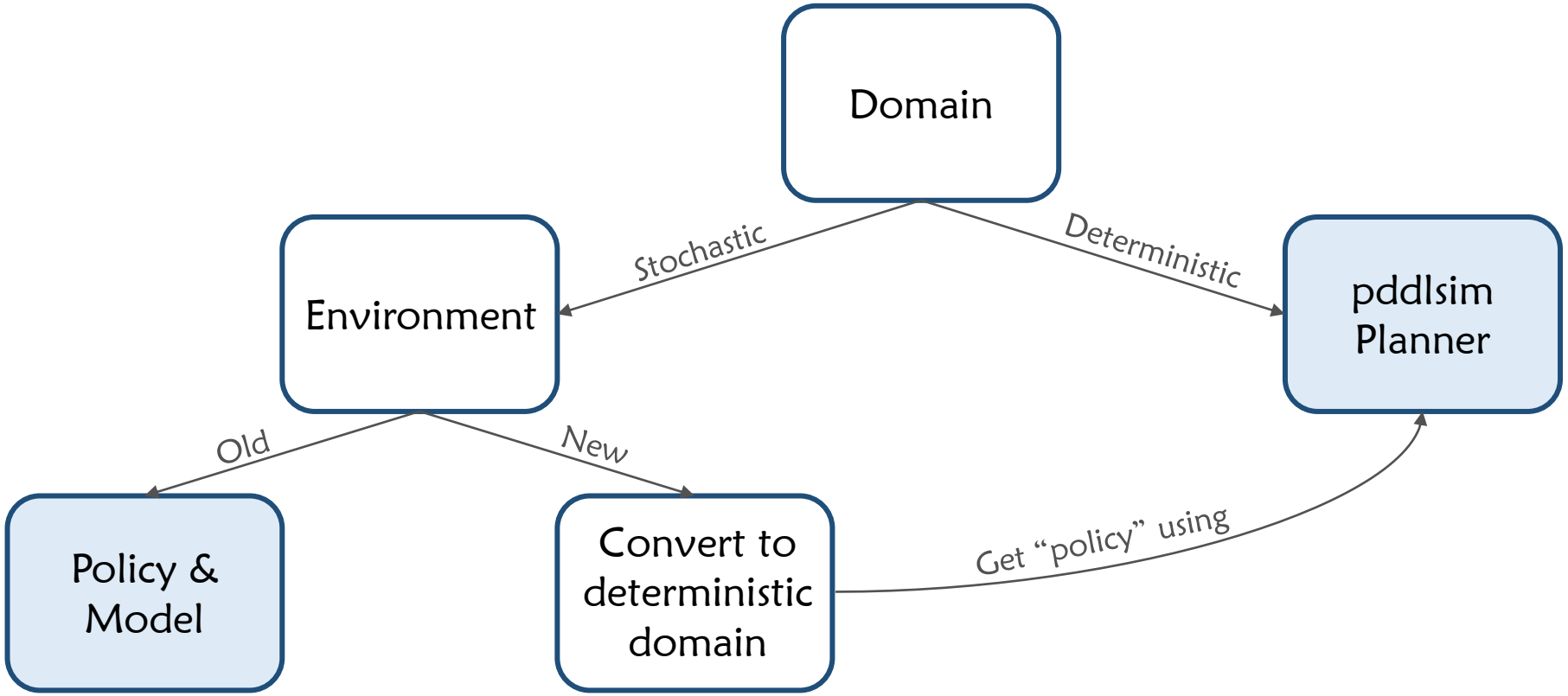


Learning Phase

- Rewards:
  - Positive reward for achieving a goal.
  - Negative reward for getting to dead end.
  - Negative reward for a regular step.

- Problem: Several sub-goals.

Execution Phase



Execution Phase

- Infinite loop handling:
  - Choose the action with the second highest Q-value.
  - Choose a random action.