LECTURER: TAI LE QUY

INTRODUCTION TO REINFORCEMENT LEARNING

UNIT 6

SUMMARY: INTRODUCTION TO REINFORCEMENT LEARNING

STUDY GOALS

- Understand core reinforcement learning concepts: states, actions, rewards and how such systems learn from experience
- Differentiate RL categories and identify methods in each
- Apply and evaluate RL methods to solve practical problems
- Evaluate and describe how deep neural networks are pushing progress in RL

EXPLAIN SIMPLY

1. Explain the fundamental components of RL and how they relate to Markov decision processes?

2. Describe the categorizations of RL methods and identify specific techniques that fall under each category

3. Compare RL methods (policy iteration, value iteration, SARSA, Q-Learning) and evaluate their effectiveness in various scenarios

WHY REINFORCEMENT LEARNING

Robotics

Gaming

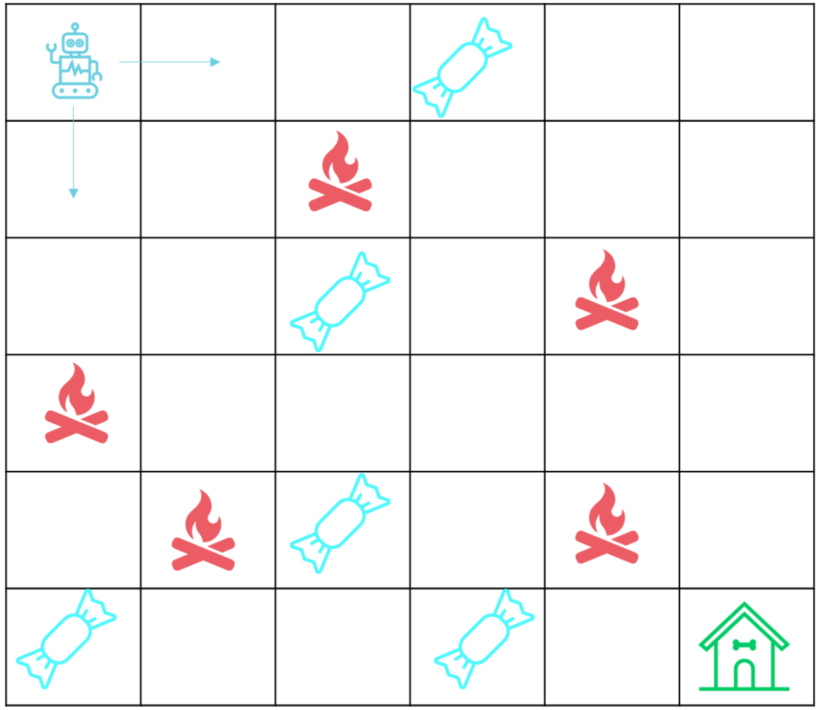
Reinforcement Learning

Autonomous vehicles

Language Processing

Objective: learn to make optimal decisions from experience

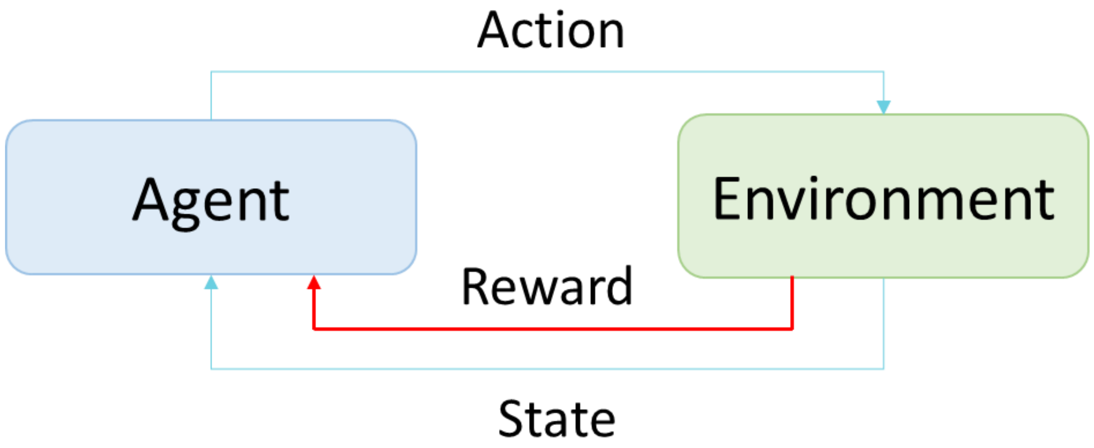
- Learn through trial and error
- Use experience to improve decisions
- Reward desired behavior
- Punish undesired behavior



Source of the image: Plaku, 2023.

Task: the agent perceives the environment and acts optimally

- Agent: learning entity;
  environment: learning context
- State: current condition of the world;
  action: decision made by agent
- Goal: optimal policy that
  maximizes rewards



Source of the image: Plaku, 2023.

MARKOV DECISION PROCESS

State

- Possible condition of system

Action

- Choices available in each state

Reward

- Feedback from environment

Transition

- Probabilities of moving between states

MDP

RL METHODS CATEGORIZATION OVERVIEW

Dynamic Programming

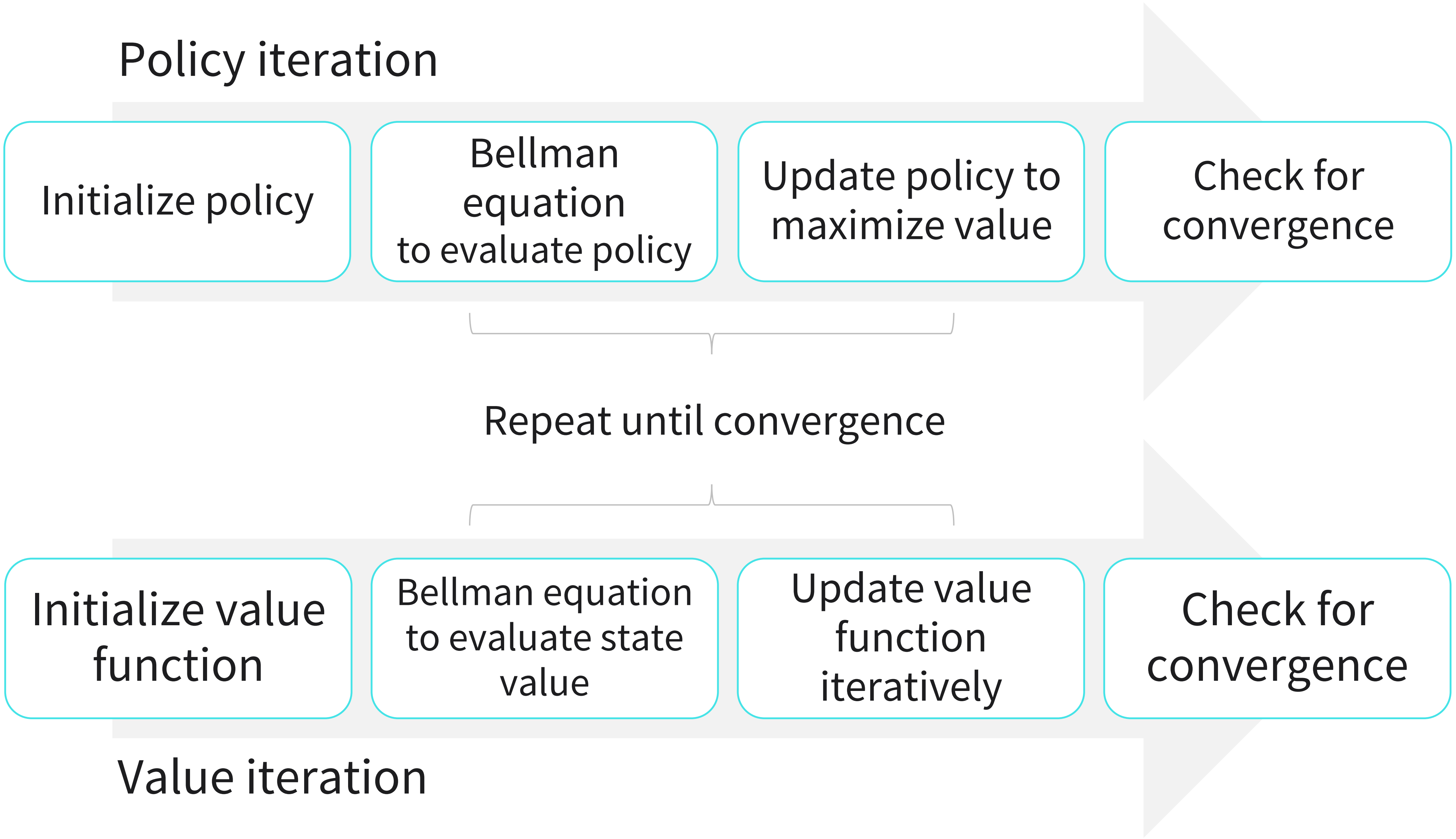
- Model based
- Compute value function iteratively
- Value iteration & policy iteration

Monte Carlo Methods

- Model free
- Use episodes to estimate values
- Handle large requests

Temporal Difference Methods

- Model free
- Bootstrapping to estimate values
- Q-learning and SARSA


Policy iteration
Initialize policy
Bellman equation to evaluate policy
Update policy to maximize value
Check for convergence
Repeat until convergence
Initialize value function
Bellman equation to evaluate state value
Update value function iteratively
Check for convergence
Value iteration

TD algorithm: efficient learning from raw experiences

Observe current state and choose action → Observe reward and new state → Calculate TD error → Update value function

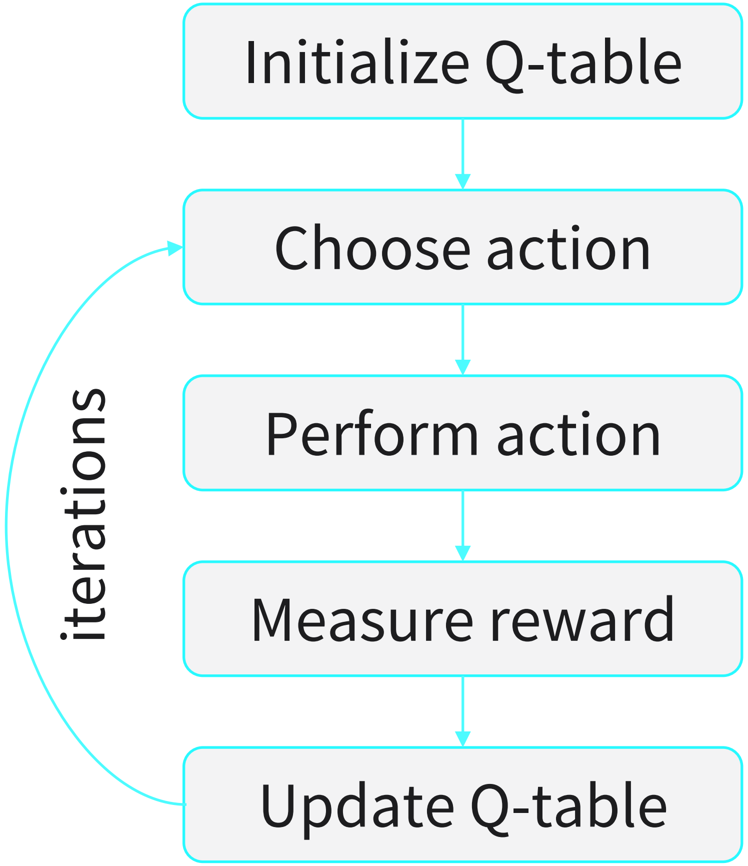
$TD(0)$

Multi-step TD

$TD(\lambda)$

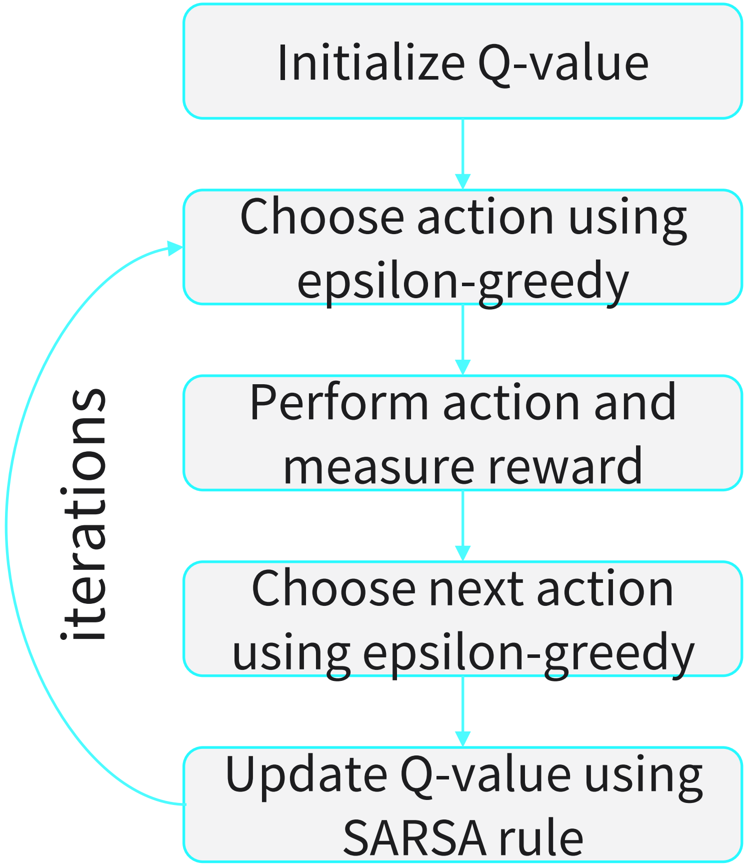
Q-learning: simple and effective RL algorithm

- Learn optimal policy through exploration and exploitation
- Update action-value function based on experience
- Maximize expected cumulative reward



Iteratively estimate Q-value using state-action pairs

- Learn optimal action-value function
- Maximize cumulative reward
- Adapt to changing environment



Synergy of neural networks and reinforcement learning

Replay buffer

- Store experience in replay buffer
- Sample mini-batch to update weights
- Improve efficiency and stabilize learning

Target network

- Non-stationary targets in RL
- Target duplicates main network
- Target lags behind a few time steps

Network training

- Agents balance exploration vs exploitation
- Combine Bellman equations with greedy-epsilon strategy

REVIEW STUDY GOALS

- Understand core reinforcement learning concepts: states, actions, rewards and how such systems learn from experience
- Differentiate RL categories and identify methods in each
- Apply and evaluate RL methods to solve practical problems
- Evaluate and describe how deep neural networks are pushing progress in RL

SESSION 5

TRANSFER TASK

Case study

Your task is to train an autonomous agent to navigate in complex cluttered environments. The agent must be able to perceive and interpret the state of the environment, including road signs, traffic lights, pedestrians, and other vehicles

Task

Model this problem as a Reinforcement Learning Task.
Discuss which would be the key components you would consider.
Which algorithms would you utilize?

Please present your results.
The results will be discussed in plenary.

1. The goal of Reinforcement Learning is
   a) Learn to act
   b) Learn to generalize
   c) Learn to specialize
   d) Learn to compress

2. Policy is formally defined as a mapping from states to which quantity?
   a) Return
   b) Rewards
   c) Actions
   d) Value function

3. SARSA and Q-learning belong to which family of methods

   a) Monte Carlo methods
   b) Dynamic programming methods
   c) Temporal difference methods
   d) Machine learning methods

LIST OF SOURCES

Images

Plaku, 2023.

How did you like the course?

HOW DID YOU
LIKE THE COURSE?