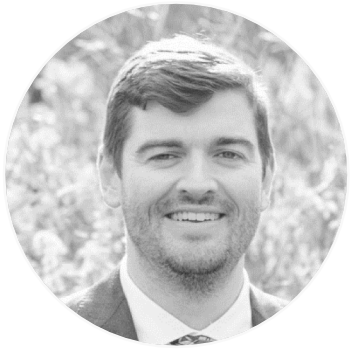
Implementing Reinforcement Systems



Justin Flett
PROFESSOR – FACULTY OF APPLIED SCIENCE

Module Overview

Understand reinforcement learning systems

Implementation of a simple reinforcement learning system

Introduce Markov decision process (MDP) and Policy-based Agents

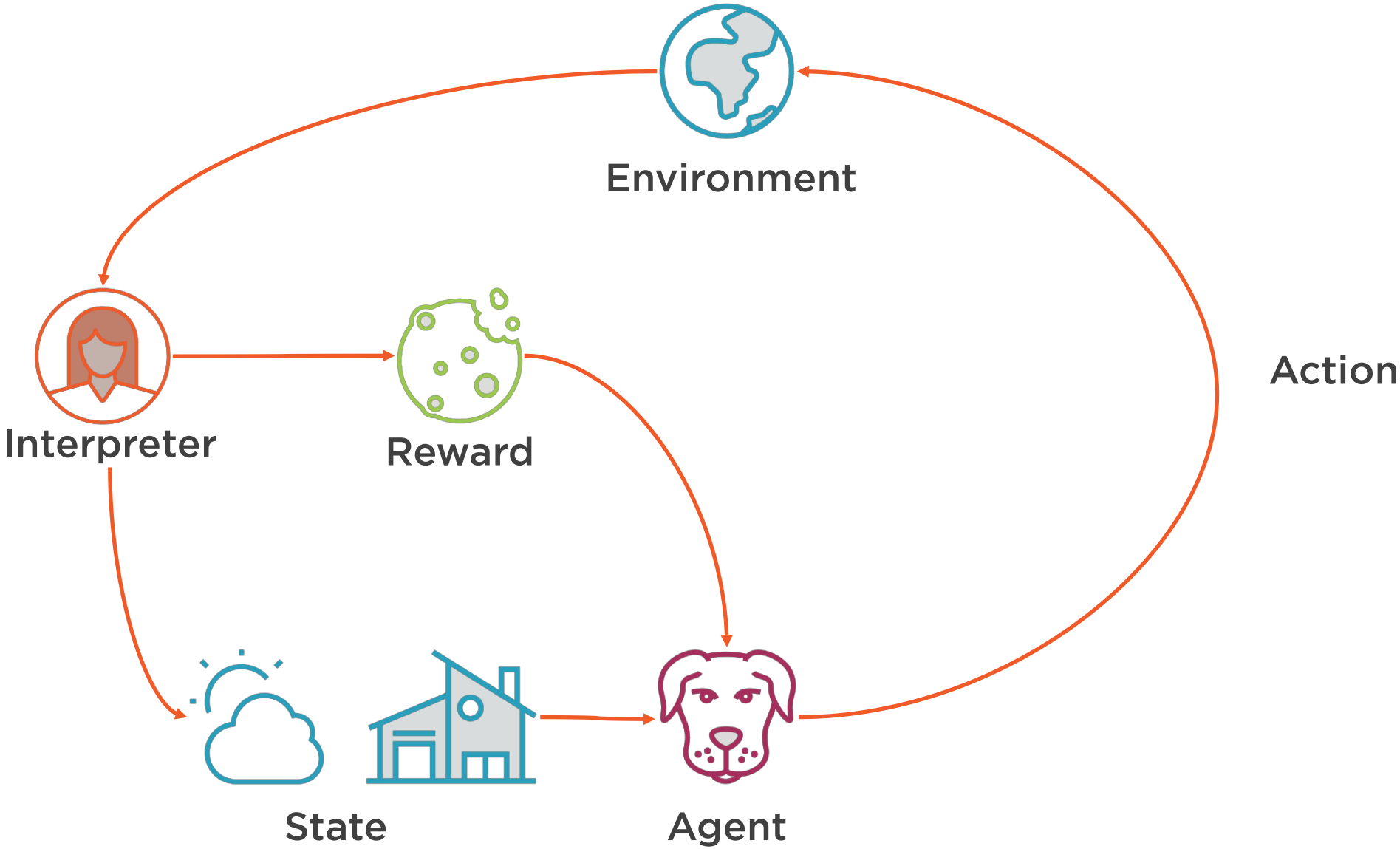
Discuss additional learning tools and next steps for further learning

Understanding Reinforcement Learning

Reinforcement Learning

A learning system concerned with how an agent should take actions in an environment in order to maximize some reward.

Reinforcement Learning



Supervised / Unsupervised vs. Reinforcement

Supervised / Unsupervised Learning

Objective is to predict, classify or simply some data

Environment is known (inputs known)

Train models by finding patterns in data

Reinforcement Learning

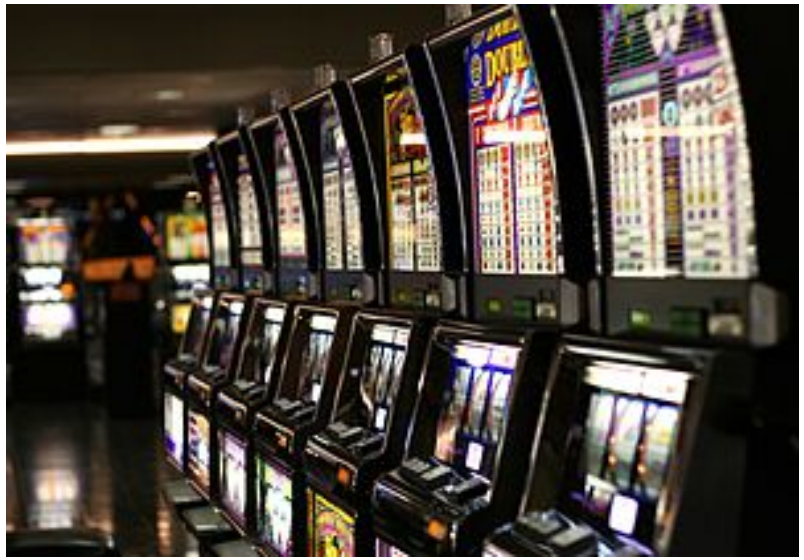
Objective is to choose the best actions that will give the best rewards

Environment is unknown

Train models by exploring the environment to determine rewards

Implementing a Simple Reinforcement Learning System

Multi-armed Bandit



Common problem: Slot machines have different winning probabilities

Each machine you can win (positive reward) or lose (negative reward)

How can you (the agent) find out which machine is best?

Policy

Defines the agents way of behaving within the environment. The mapping from of an environment including all actions and states.

Policy Gradient Method

Reinforcement learning technique that optimizes a policy using gradient descent methods.

Epsilon-greedy Policy

The best lever is selected for a proportion of $(1 - \varepsilon)$ of the trials, and a lever is selected at random for a proportion ε.

Understanding Markov Decision Process and Reinforcement Policies

Markov Decision Process (MDP)

A process for modeling your full environment where the outcomes are partly random and partly under the control of a decision maker.

Markov Decision Process

Set of states	S

Markov Decision Process

Set of states	S
Set of actions	A

Markov Decision Process

Set of states	S
Set of actions	A
Probability of transition from state s to s'	$P_a(s, s') = \mathrm{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$

Markov Decision Process

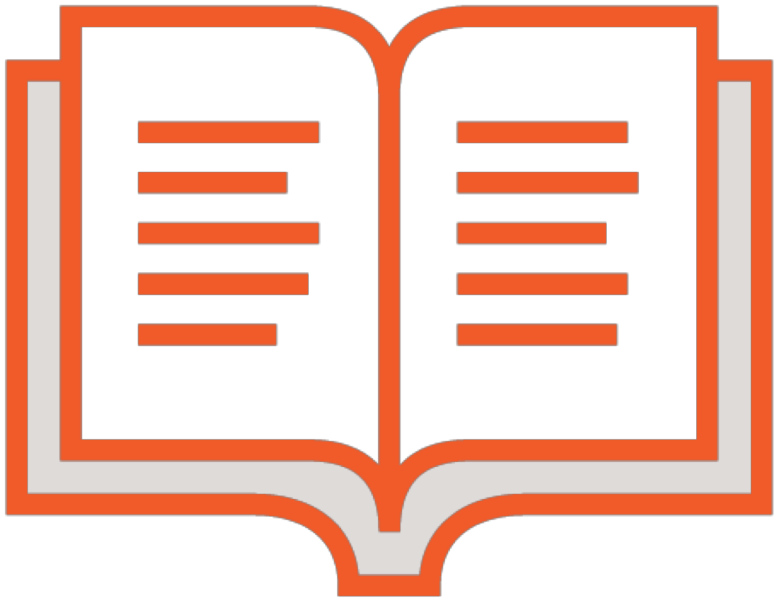
Set of states	S
Set of actions	A
Probability of transition from state s to s'	$P_a(s, s') = \mathrm{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$
Reward	$R_a(s, s')$

Markov Decision Process

Set of states	S
Set of actions	A
Probability of transition from state s to s'	$P_a(s, s') = \mathrm{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$
Reward	$R_a(s, s')$
Discount factor	$\gamma \in [0,1)$

Further Learning and Next Steps

Further Learning



Tensorflow website (tensorflow.org)

- Tutorials, videos, documentation

Pluralsight courses

- Building Classification Models with TensorFlow
- Understanding Algorithms for Reinforcement Learning

OpenAI gym toolkit for reinforcement learning environments

Summary

- Understood and implemented supervised learning systems
- Implemented a movie recommendation learning system
- Understood and implemented a simple reinforcement learning system
- Discussed additional learning tools and next steps for further learning