СВ

Aditya, Abhinandar

Introduction

Methods Algorithms General

Observations

Observations

Q-Learning DP vs Q-Learn

Future Work

References

# Change in Behavior and Transfer with Change in Properties of the Environment

Aditya Anantwar Abhinandan De

Instructor: Prof. Aritra Hazra
IIT Kharagpur

Term Project September 18, 2022

#### **Presentation Overview**

СВ

Aditya, Abhinandar

Introduction

Methods Algorithms General Methodology

Observations

DP

Q-Learning

DP vs Q-Learning

Future Work

1 Introduction

2 Methods Algorithms General Methodology

3 Observations
DP
Q-Learning
DP vs Q-Learning
Discussion

4 Future Work

## Introduction

СВ

Aditya, Abhinandan

Introduction

Methods Algorithms General Methodology

Observations

Q-Learning
DP vs Q-Learnin

**Future Work** 

References

- Aim of project is to analyze change in behaviour of RL based systems with change in properties.
- We utilize OpenAl's CartPole environment [1] for analysis.

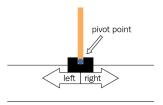


Figure: The CartPole problem.

CE

Aditya, Abhinandar

Introduction

Methods Algorithms General Methodology

DP Q-Learning DP vs Q-Learning Discussion

Future Work

#### **Dynamic Programming**

- Model based algorithm [2]
- Bellman backups

$$V_* = \max_{a} \mathbb{E}[R_{t+1} + \gamma V_*(S_{t+1}) | S_t = s, A_t = a]$$

$$Q_*(s, a) = \mathbb{E}[R_{t+1} + \gamma \max_{a'} Q_*(S_{t+1}, a') | S_t = s, A_t = a]$$

#### **Q-Learning**

- Model free algorithm [2]
- State action pair backups

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_{a} Q(S_{t+1}, a) - Q(S_t, A_t)]$$

## **General Method**

CB

Aditya, Abhinandar

Introduction

Algorithms

General Methodology

Observatior

DP Odlearning

DP vs Q-Learn

Future Work

References

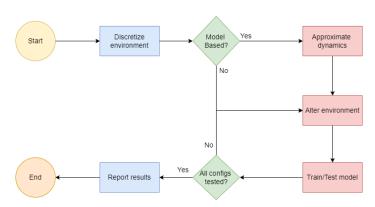


Figure: Flowchart describing our process.

## **DP Observations**

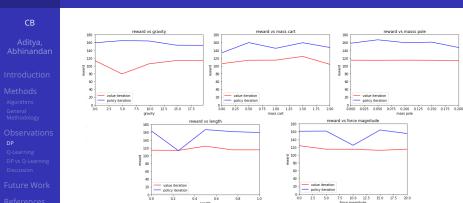


Figure: Average rewards per 1000 episodes for testing under the 25 different scenarios. The stable nature of the graphs must be noted as DP is known for producing a stable policy with lower variance.

## Q-Learning Observations

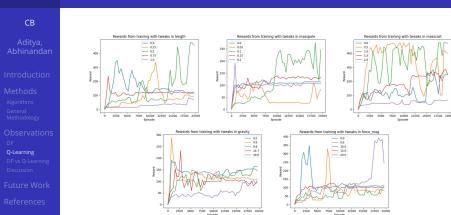


Figure: Rewards obtained while training the Q-Learning algorithm for different configurations. One can observe the variable convergence and high variance in the rewards obtained.

## Policy Iterations vs Q-Learning

Aditya,
Abhinandan

Introduction

Wethods
Algorithms
General
Methodology

Disservations
DP
QQLearning
DP vs Q-Learning
Discussion

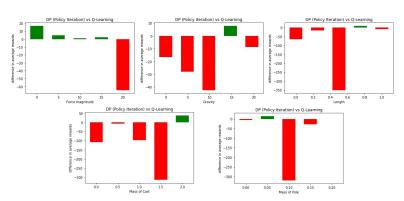


Figure: Policy Iteration vs Q-Learning: difference between rewards. A red bar indicates that Q-Learning performs better than Policy Iteration while a green bar indicates otherwise.

## Value Iterations vs Q-Learning

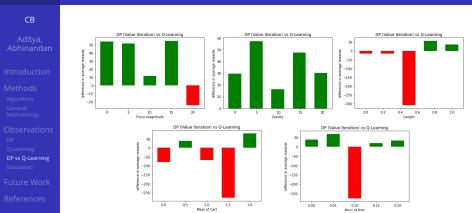


Figure: Value Iteration vs Q-Learning: difference between rewards. A red bar indicates that Q-Learning performs better than Value Iteration while a green bar indicates otherwise.

## Discussion

СВ

Aditya, Abhinandan

Methods
Algorithms

Observations
DP
Q-Learning
DP vs Q-Learning
Discussion

Future Work References

- There exists a bias variance tradeoff between DP and Q-Learning.
- The performance of DP is much more stable as compared to Q-Learning. The latter behaves in a very unpredictable manner.
- The variance of Q-Learning my be seen via the following table

	Mean	Standard deviation
Policy Iteration	114.826	6.021
Value Iteration	152.147	15.477
Q-Learning	276.728	149.415

Figure: Table for comparing mean and variance for the unaltered environment case.

#### **Future** work

СВ

Aditya, Abhinanda

Introduction

Methods Algorithms General Methodolog

Observation

DP

Q-Learning

DP vs Q-Learning

Discussion

Future Work

 Extension to value function approximation to tackle the continuous state space better.

- Usage of better convergence techniques by altering the learning rate formulation.
- Exploration of transfer learning for better generalization with change in configuration.

## References

СВ

Aditya, Abhinanda

Introduction

Methods
Algorithms
General
Methodolog

Observations

DP Q-Learning DP vs Q-Learning Discussion

Future Worl

References

[1] Greg Brockman et al. OpenAl Gym. 2016. eprint: arXiv:1606.01540.

[2] Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. Second. The MIT Press, 2018. URL: http://incompleteideas.net/book/the-book-2nd.html.

4 D > 4 B > 4 B > 4 B > 9 Q P

## The End

C

Aditya, Abhinandai

Introduction

Methods Algorithms

Methodology

Observations

Q-Learning DP vs Q-Learn

Future Work

References

## Thank You!