Dynamic Programming and Reinforcement Learning

Introduction (Week 1)

Rainer Schlosser, Alexander Kastius

Hasso Plattner Institute (EPIC)

April 21, 2022

The World is Full of Decision Problems

BOARDING PASS Priority 3 AIR

FLIGHT 125 to Munich

Terminal 3 Boarding Area C

PASSENGER Lorem Ipsum

DEPARTURE August 21, 2016

SEAT 17B

Economy DEPART 12:34 PM ARRIVE 6:50 AM

FLIGHT 125 to Munich

T 3 Area C

PASSENGER Lorem Ipsum

Economy

BOARDING PASS Priority 1 AIR

FLIGHT 125 to Munich

Terminal 3 Boarding Area C

PASSENGER Lorem Ipsum

DEPARTURE August 21, 2016

SEAT 2B

First Class DEPART 12:34 PM ARRIVE 6:50 AM

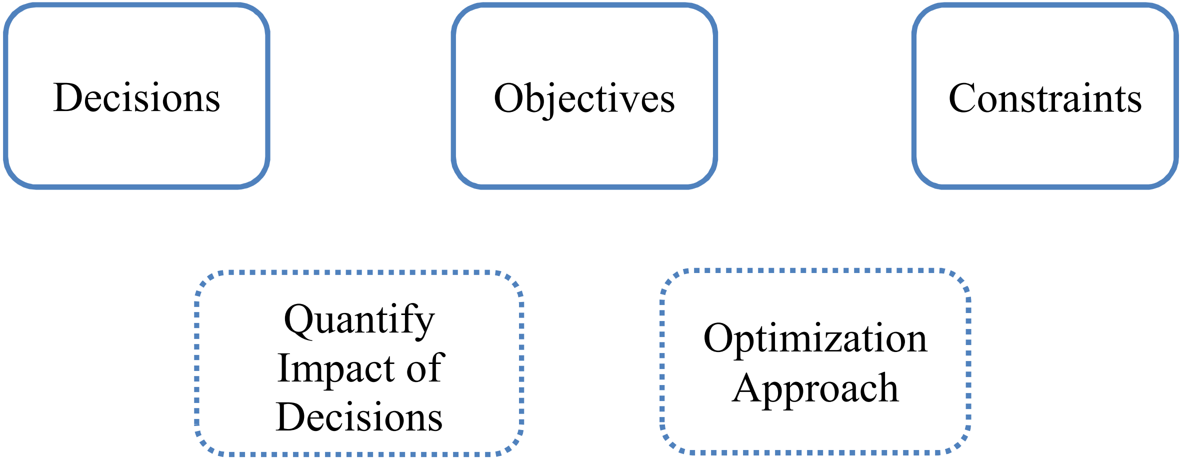
FLIGHT 125 to Munich

T 3 Area C

PASSENGER Lorem Ipsum

First Class

How to Approach Decision Problems?



Decisions

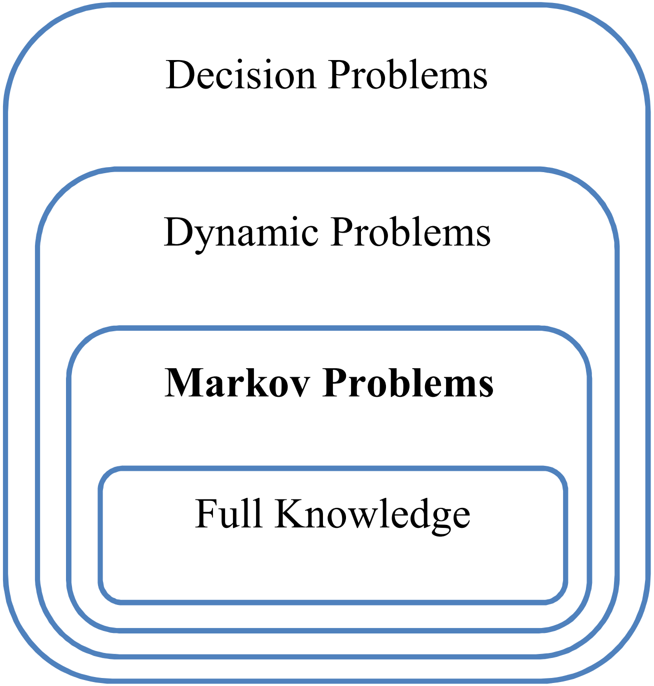
Objectives

Constraints

Quantify Impact of Decisions

Optimization Approach

Dynamic Problems & Markov Decision Processes (MDP)

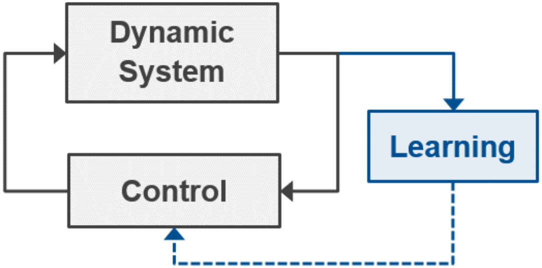


Decision Problems

Dynamic Problems

Markov Problems

Full Knowledge



Types of Decision Problems

Static vs. **Dynamic**

Deterministic vs. **Stochastic** vs. Uncertainty

Low vs. **Large** Complexity/ Dimensionality

Finite vs. **Infinite** Time

Discrete vs. Continuous Time

Discrete vs. Continuous Actions & Rewards

Risk Neutral vs. Risk-Averse

Markov Type (**yes** / no)

System Dynamics **Known/Unknown**

System Fully vs. Partially Observable

Further Players/Agents

Solution Methods

Optimal vs. Heuristic Solutions

Optimal

- (Exhaustive Search)
- Dynamic Programming

Heuristic Solutions

- Approximate Dynamic Programming
- Reinforcement Learning
  - Q-Learning
  - Deep Q-Learning
  - Policy Gradient Algorithms

Agenda

- Introduction ✓
- **Personal Background**
- Structure of the Course & Grading

Personal Background (Rainer)

- Ph.D. Operations Research (2014), Humboldt-University of Berlin
- Hasso Plattner Institute (EPIC) since 2015
- Field of Research
  - Data-driven decision support
  - Focus on stochastic models & Dynamic Programming (DP)
- Current Areas of Applications
  - Revenue management (e.g., dynamic pricing, ordering, advertising, risk)
  - Database configuration (e.g., data placement problems, index selection)

Personal Background (Alex)

- Master Computer Science (2020), HPI
- Hasso Plattner Institute (EPIC), PhD Candidate
- Field of Research
  - Data-driven decision support
  - Focus on Reinforcement Learning (RL)
- Current Areas of Applications
  - Revenue management
  - RL methods

What about you?

- Background?
- Interests?
- Expectations?
- Questions?
- Online vs. Offline?

Agenda

- Introduction ✓
- Personal background ✓
- **Structure of the Course & Grading**

Technical Information

Credits?	4 SWS (V/Ü), 6 ECTS (graded)		
When?	Monday	15.15 – 16.45	VL/UE (lecture/exercise)
	Thursday	13.30 – 15.00	VL/UE (lecture/exercise)
	Start:	April 21, 2022	End: July 25, 2022
Where?	currently via Zoom	(Room: 7271364393, Password: 256757)	
Who?	Rainer Schlosser, rainer.schlosser@hpi.de Alexander Kastius, alexander.kastius@hpi.de		
Slides?	EPIC, Teaching, Summer 2022		

Structure of the Course

- April/May: Lectures on models & methods:
  - (i) Markov Decision Processes (MDPs)
  - (ii) Dynamic Programming (DP)
  - (iii) Reinforcement Learning (RL)

- June/July: Choose projects, apply/extend suitable techniques

  Work in teams, input/support will be given

- July/Aug: Documentation of projects results

Goals of the Course & Grading

- Goal: Develop models to compute optimized decisions for different problems & applications
- Learn: Optimization techniques
- Do: Apply & extend different optimization approaches
- Grading:
  - (i) Presentation of project results (~July)
  - (ii) Documentation ("Projektarbeit")
    Deadline Sep 15 (~10-20 pages)

Prerequisites

- Programming

  Parameters, Data preparation
  Loops, Recursions, Simulations

- Basic Mathematical Background

  Sets, Vectors
  Probabilities, Random variables, Expected values

- More does not harm

  Regression analysis
  NNs
  Deep learning
  Game theory

Overview

Week	Dates	Topic	
1	April 21	Introduction	
2	April 25/28	Finite + Infinite Time MDPs	
3	May 2/5	Dynamic Programming (DP) Exercise	
4	May 9/12	Approximate Dynamic Programming (ADP) + Q-Learning (QL)	
5	May 16/19	Deep Q-Networks (DQN)	
6	May 23	DQN Extensions	(Thu May 26 “Himmelfahrt”)
7	May 30/June 2	Policy Gradient Algorithms	
8	June 9	Project Assignments	(Mon June 6 “Pfingstmontag”)
9	June 13/16	Work on Projects: Input/Support	
10	June 20/23	Work on Projects: Input/Support	
11	June 27/30	Work on Projects: Input/Support	
12	July 4/7	Work on Projects: Input/Support	
13	July 11/14	Work on Projects: Input/Support	
14	July 18/21	Final Presentations	
	Sep 15	Finish Documentation	

What are Dynamic Optimization Problems?

- How to control a dynamic system over time?

- Instead of a single static decision we have a *sequence* of decisions

- The system evolves over time according to a certain dynamic

- The decisions are supposed to be chosen such that

  a certain objective/quantity/criteria is optimized

- Find the right balance between short and long-term effects

Examples Please!

Examples

- Inventory Replenishment
- Selling Airline Tickets
- Drinking at a Party
- Exam Preparation
- Brand Advertising
- Used Cars
- Eating Cake

Task: Describe & Classify

- Goal/Objective
- State of the System
- Actions
- Dynamic of the System
- Revenues/Costs (Rewards)
- Finite/Infinite Horizon
- Stochastic Components

Classification

Example	Objective	State	Action	Rewards	Dynamic
Inventory Mgmt.	min costs	#items	#order	order/holding	entry-sales

Classification

Example	Objective	State	Action	Rewards	Dynamic
Inventory Mgmt.	min costs	#items	#order	order/holding	entry-sales
Airline Tickets					
Drinking at Party					
Exam Preparation					
Advertising					
Used Cars					
Eating Cake					

Classification

Example	**Objective**	**State**	**Action**	**Rewards**	**Dynamic**
Inventory Mgmt.	min costs	#items	#order	order/holding	entry-sales
Airline Tickets	max revenue	#tickets	#price	sales	current-sold
Drinking at Party	max fun	‰	#beer	fun/money	impact-rehab

Classification

Example	Objective	State	Action	Rewards	Dynamic
Inventory Mgmt.	min costs	#items	#order	order/holding	entry-sales
Airline Tickets	max revenue	#tickets	#price	sales	current-sold
Drinking at Party	max fun	‰	#beer	fun/money	impact-rehab
Exam Preparation	max mark/effort	#learned	#learn	effort, mark	learn-forget
Advertising	max profits	image	#advertise	campaigns	effect-forget

Classification

Example	**Objective**	**State**	**Action**	**Rewards**	**Dynamic**
Inventory Mgmt.	min costs	#items	#order	order/holding	entry-sales
Airline Tickets*	max revenue	#tickets	#price	sales	current-sold
Drinking at Party*	max fun	‰	#beer	fun/money	impact-rehab
Exam Preparation*	max mark/effort	#learned	#learn	effort, mark	learn-forget
Advertising	max profits	image	#advertise	campaigns	effect-forget
Used Cars	min costs	age	replace(y/n)	buy/repair	aging/faults
Eating Cake*	max utility	%cake	#eat	utility	outflow

* Finite horizon

General Problem Components

- What do you want to optimize (e.g., expected rewards) (Objective)
- Define the state of your system (State)
- Define the set of possible actions (state dependent) (Actions)
- Quantify event probabilities (state+action dependent) (Dynamics) (!)
- Define rewards (state+action+event dependent) (Rewards)
- Define state transitions (state+action+event dependent) (Transitions)
- What happens at the end (of the time horizon)? (Final Rewards)

Recall - Questions?

- Finite/Infinite Time Horizon
- States
- Actions
- Events & Rewards
- Dynamics & State Transitions
- Deterministic/Stochastic
- Discrete/Continuous

Overview

Week	Dates	Topic	
1	April 21	Introduction	
2	April 25/28	Finite + Infinite Time MDPs	
3	May 2/5	Dynamic Programming (DP) Exercise	
4	May 9/12	Approximate Dynamic Programming (ADP) + Q-Learning (QL)	
5	May 16/19	Deep Q-Networks (DQN)	
6	May 23	DQN Extensions	(Thu May 26 “Himmelfahrt”)
7	May 30/June 2	Policy Gradient Algorithms	
8	June 9	Project Assignments	(Mon June 6 “Pfingstmontag”)
9	June 13/16	Work on Projects: Input/Support	
10	June 20/23	Work on Projects: Input/Support	
11	June 27/30	Work on Projects: Input/Support	
12	July 4/7	Work on Projects: Input/Support	
13	July 11/14	Work on Projects: Input/Support	
14	July 18/21	Final Presentations	
	Sep 15	Finish Documentation	