Lecture 1a: Introduction
CS885 Reinforcement Learning

2025-01-07

Complementary readings: [SutBar] Chapter 1, [Sze] Chapter 1

Pascal Poupart
David R. Cheriton School of Computer Science

UNIVERSITY OF WATERLOO

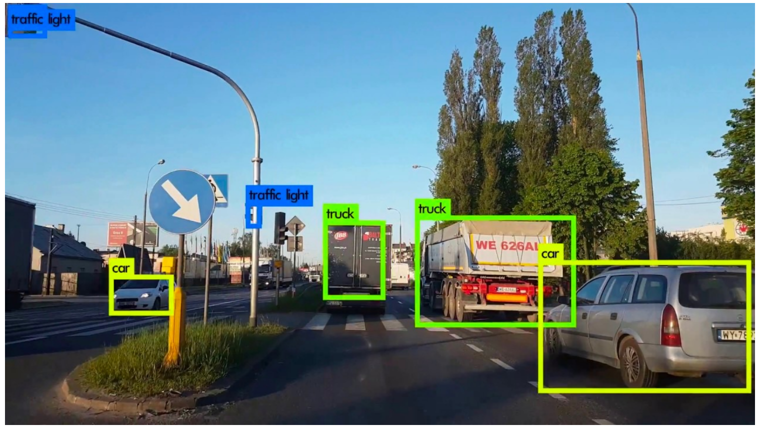
Outline

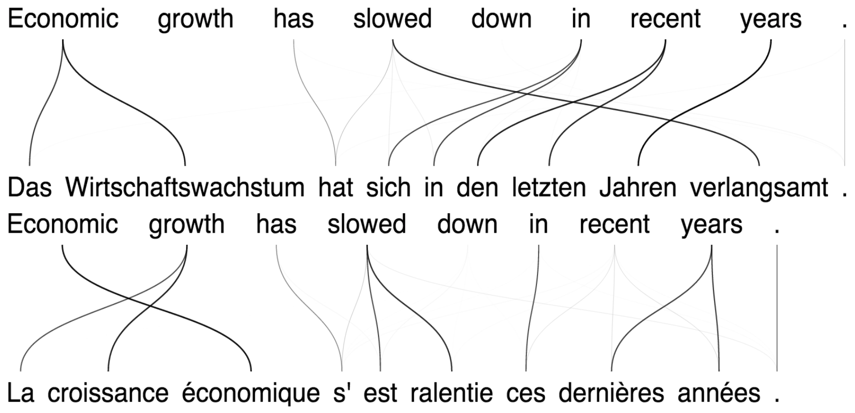
- Introduction to Reinforcement Learning
- Course logistics

Machine Learning

- Traditional computer science
  - Program computer for every task

- New paradigm
  - Provide examples to machine
  - Machine learns to accomplish tasks based on examples





Economic growth has slowed down in recent years .

Das Wirtschaftswachstum hat sich in den letzten Jahren verlangsamt .

Economic growth has slowed down in recent years .

La croissance économique s' est ralentie ces dernières années .

Machine Learning

- Success mostly due to supervised learning
  - Bottleneck: need lots of labeled data
  - Limitation: mimic data

- Alternatives
  - Unsupervised, semi-supervised, self-supervised learning
  - Transfer learning, domain adaptation, meta-learning
  - Reinforcement Learning

What is Reinforcement Learning?

- Reinforcement learning is also known as
  - Optimal control
  - Approximate dynamic programming
  - Neuro-dynamic programming

- Wikipedia: reinforcement learning is an area of machine learning inspired by behavioural psychology, concerned with how software **agents** ought to take **actions** in an **environment** so as to maximize some notion of cumulative **reward**.

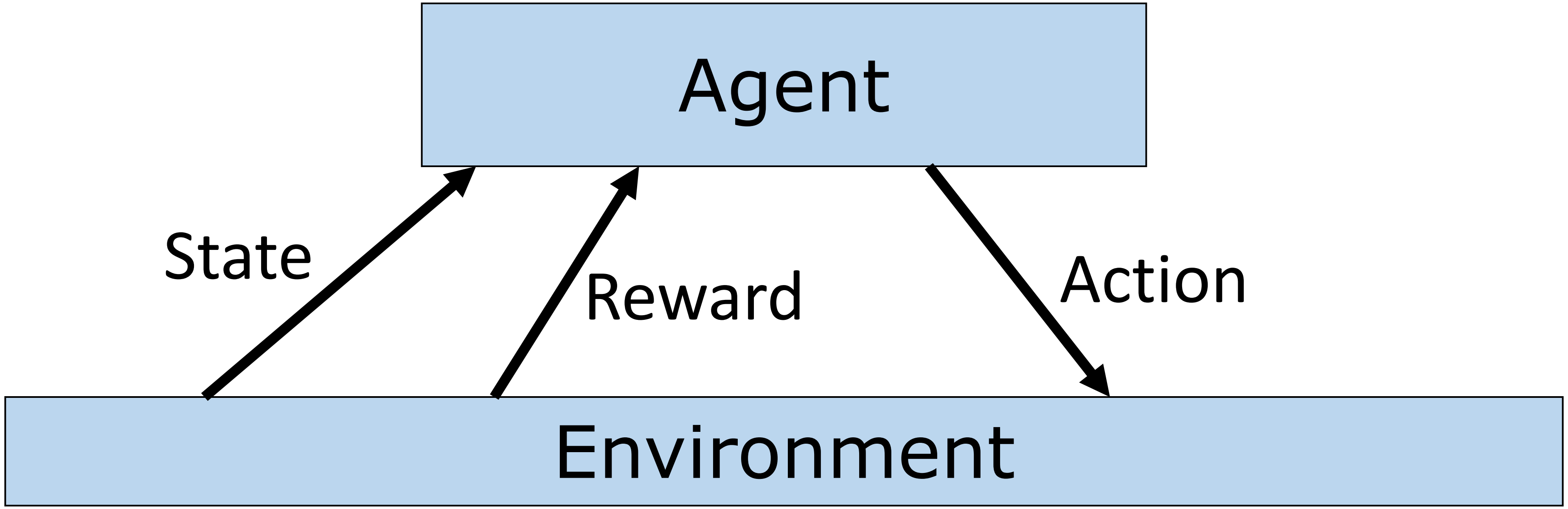
Animal Psychology

- Negative reinforcements
  - Pain and hunger
- Positive reinforcements
  - Pleasure and food

- Reinforcements used to train animals

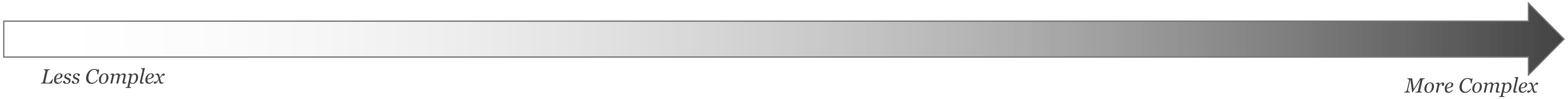
- Let’s do the same with computers

Reinforcement Problem



Goal: Learn to choose actions that maximize rewards

Sample Industrial Use Cases


Less Complex

More Complex

Contextual Bandits

Marketing

ad placement,
recommender systems

Loyalty programs

personalized offers

Price management

airline seat pricing
cargo shipment pricing
food pricing

Optimal design

interface personalization

Bayesian Optimization

Hyperparameter optimization

Troubleshooting

Customer assistance

Diagnostics

Fault detection

Design of experiments

Drug design
Material design

Sequential decision Making

Automated trading

Stocks, energy

Optimization

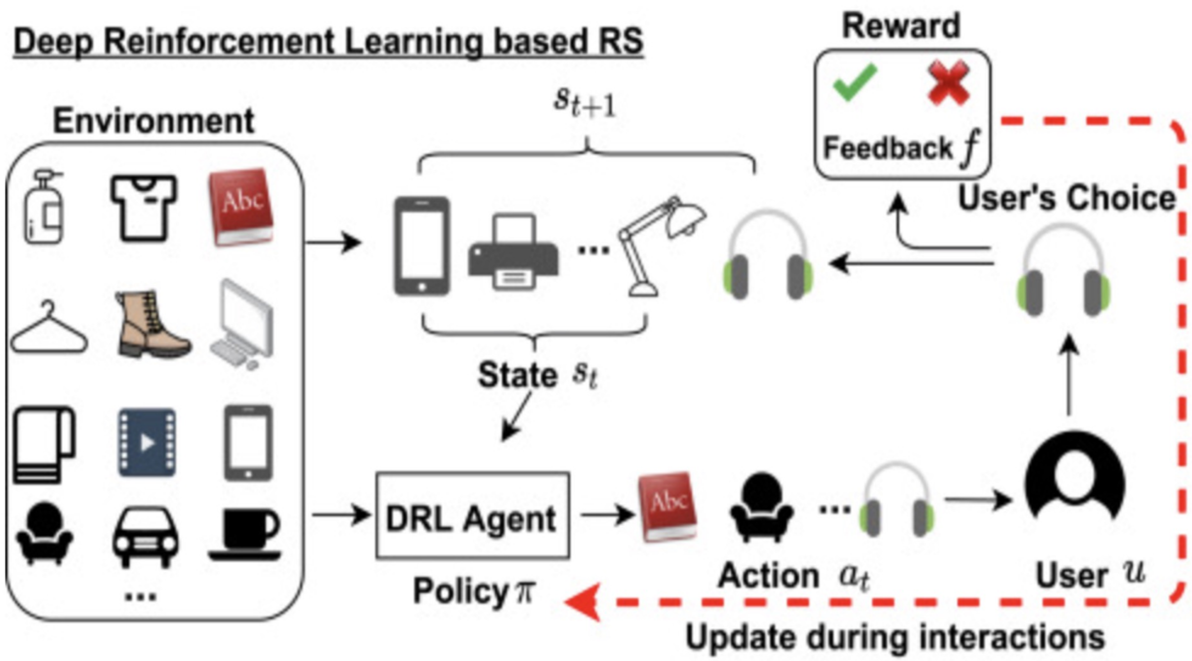
Path planning
Routing
Energy consumption

Control

Robotics
Autonomous driving

Marketing (Recommender System)

- **Agent:** recommender system
- **Environment:** user
- **State:** context, past recommendations and feedback
- **Action:** recommended item
- **Reward:** value of user feedback

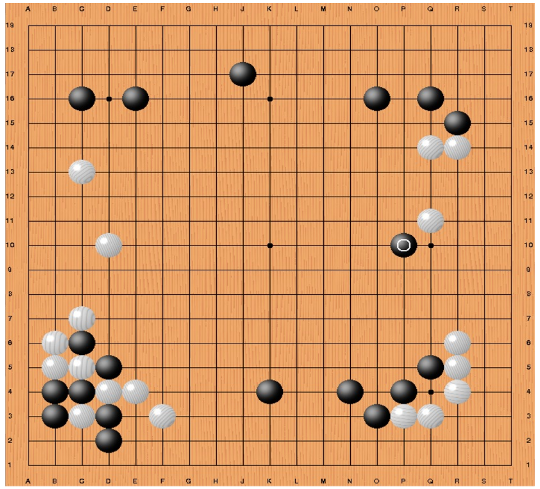


Operations Research (vehicle routing)

- **Agent:** vehicle routing system
- **Environment:** stochastic demand
- **State:** vehicle location, capacity and depot requests
- **Action:** vehicle route
- **Reward:** - travel costs

Game Playing (Computer Go)

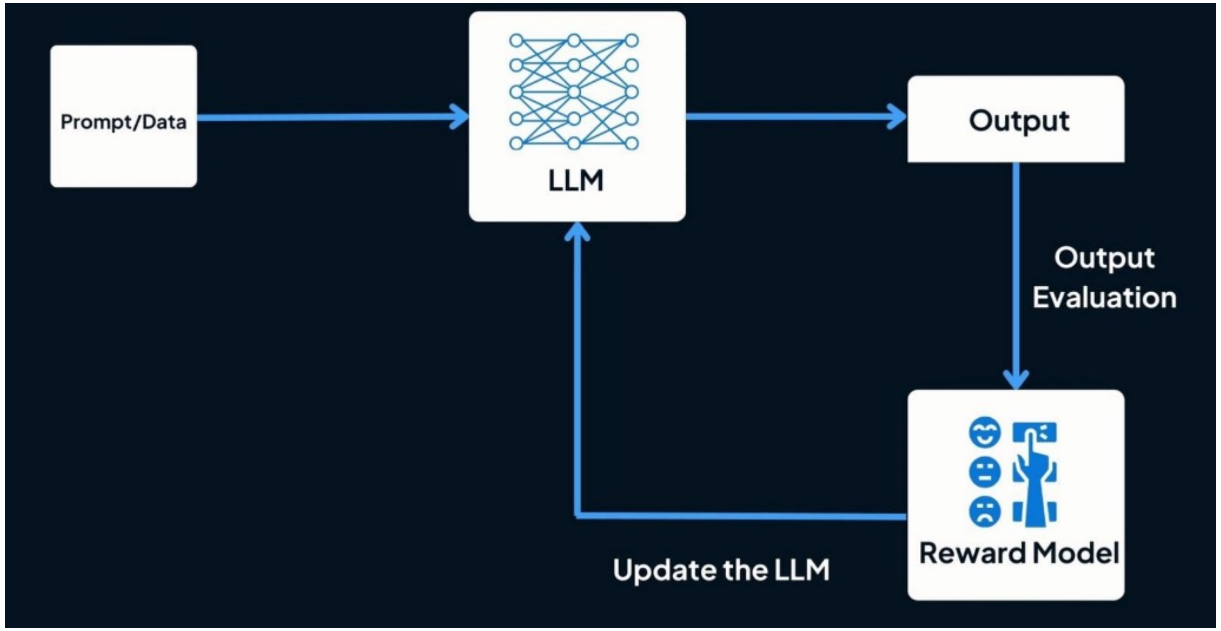
- **Agent:** player
- **Environment:** opponent
- **State:** board configuration
- **Action:** next stone location
- **Reward:** +1 win / -1 loose



- 2016: AlphaGo defeats Lee Sedol (4-1)
  - Game 2 move 37: AlphaGo plays unexpected move (odds 1/10,000)

Large Language Model (RL from Human Feedback)

- **Agent:** system
- **Environment:** user
- **State:** history of past utterances
- **Action:** system utterance
- **Reward:** task completion, human feedback



Credit: https://www.twine.net/blog/what-is-reinforcement-learning-from-human-feedback-rlhf-and-how-does-it-work/

Computational Finance (Trading)

- **Agent:** trading software
- **Environment:** other traders
- **State:** price history
- **Action:** buy/sell/hold
- **Reward:** amount of profit

Example: how to purchase a large # of shares in a short period of time without affecting the price

Reinforcement Learning

- Comprehensive, but challenging form of machine learning
  - Stochastic environment
  - Incomplete model
  - Interdependent sequence of decisions
  - No supervision
  - Partial and delayed feedback

- **Long term goal**: continual machine learning