CS 747: Foundations of Intelligent and Learning Agents
CS 747: Foundations of Intelligent and Learning Agents

Agent

Environment/Task

Algorithm for...

Stock trading

Autopilot program

Airplane

You

Bicycle

Academic programme

at IIT Bombay

AlphaGo

Go

Shivaram Kalyanakrishnan (2023)

CS 747, Autumn 2023

2 / 5
CS 747: Foundations of Intelligent and Learning Agents

Agent
Algorithm for . . .

Environment/Task
Stock trading

AGENT
Think
Sense
Act

ENVIRONMENT

state reward action
CS 747: Foundations of Intelligent and Learning Agents

Agent
Algorithm for . . .

Environment/Task
Stock trading

Agent
Think

Environment/Task
Autopilot program

Environment/Task
Airplane

Agent
Sense

Envelope/Task
AlphaGo

Agent
Act

Envelope/Task
Bicycle

Envelope/Task
Academic programme at IIT Bombay
CS 747: Foundations of Intelligent and Learning Agents

Agent
Algorithm for ...
Autopilot program
You

Environment/Task
Stock trading
Airplane
Bicycle

Agent
Think
Sense
Act

Environment
state reward action

Type in your answer here
CS 747: Foundations of Intelligent and Learning Agents

Agent
Algorithm for . . .

Environment/Task
Stock trading

Agent
Autopilot program

Environment/Task
Airplane

Agent
You

Environment/Task
Bicycle

Agent
You

Environment/Task
Academic programme at IIT Bombay
CS 747: Foundations of Intelligent and Learning Agents

Agent

Algorithm for . . .

Autopilot program

You

You

AlphaGo

Environment/Task

Stock trading

Airplane

Bicycle

Academic programme at IIT Bombay

Go
Multi-armed bandits

Markov Decision Problems

Reinforcement learning

Multi-agent systems/learning
Multi-armed bandits
  ▶ The “explore or exploit” tradeoff.

Markov Decision Problems

Reinforcement learning

Multi-agent systems/learning
Multi-armed bandits
  - The “explore or exploit” tradeoff.
Markov Decision Problems
  - Sequential decision making.
Reinforcement learning

Multi-agent systems/learning
Multi-armed bandits
  ▶ The “explore or exploit” tradeoff.

Markov Decision Problems
  ▶ Sequential decision making.

Reinforcement learning
  ▶ Learning by trial and error, reward and punishment, to optimise long-term gain.

Multi-agent systems/learning
Multi-armed bandits
- The “explore or exploit” tradeoff.

Markov Decision Problems
- Sequential decision making.

Reinforcement learning
- Learning by trial and error, reward and punishment, to optimise long-term gain.

Multi-agent systems/learning
- Decision making in the presence of other decision-makers.
Multi-armed bandits
  ▶ The “explore or exploit” tradeoff.

Markov Decision Problems
  ▶ Sequential decision making.

Reinforcement learning
  ▶ Learning by trial and error, reward and punishment, to optimise long-term gain.

Multi-agent systems/learning
  ▶ Decision making in the presence of other decision-makers.

Several applications: game playing, robotics and control, planning and scheduling, on-line advertising, autonomous navigation, chemistry!
CS 747: Foundations of Intelligent and Learning Agents

- Operations Research (Dynamic Programming)
- Control Theory
- Neuroscience
- Reinforcement Learning
- Artificial Intelligence and Computer Science
- Psychology (Animal Behaviour)
**Prerequisites**: Probability, Algorithms, Programming.
**CS 747: Foundations of Intelligent and Learning Agents**

- **Prerequisites**: Probability, Algorithms, Programming.
- **Coming up in two weeks!**:

\[
\begin{align*}
B & \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^t \sum_{y=1}^t \Pr \left\{ \hat{p}_a(x) + \sqrt{\frac{2}{x} \ln(t)} \geq \hat{p}_*(y) + \sqrt{\frac{2}{y} \ln(t)} \right\} \\
& \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^t \sum_{y=1}^t \left( \Pr \left\{ \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2} \right\} + \Pr \left\{ \hat{p}_*(y) < p_* - \sqrt{\frac{2}{y} \ln(t)} \right\} \right) \\
& \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^t \sum_{y=1}^t \left( e^{-2x \left( \frac{\Delta_a}{2} \right)^2} + e^{-2y \left( \sqrt{\frac{2}{y} \ln(t)} \right)^2} \right) \\
& \leq \sum_{t=0}^{T-1} \sum_{x=\bar{u}_a^T}^t \sum_{y=1}^t \left( e^{-4 \ln(t)} + e^{-4 \ln(t)} \right) \leq \sum_{t=0}^{T-1} t^2 \left( \frac{2}{t^4} \right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.
\end{align*}
\]
**Prerequisites**: Probability, Algorithms, Programming.

Coming up in two weeks!

\[
B \leq \sum_{t=0}^{T-1} \sum_{x=\bar{a}_T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{ \hat{p}_a(x) + \sqrt{\frac{2}{x} \ln(t)} \geq \hat{p}_*(y) + \sqrt{\frac{2}{y} \ln(t)} \right\}
\]

\[
\leq \sum_{t=0}^{T-1} \sum_{x=\bar{a}_T}^{t} \sum_{y=1}^{t} \left( \mathbb{P}\left\{ \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2} \right\} + \mathbb{P}\left\{ \hat{p}_*(y) < p_* - \sqrt{\frac{2}{y} \ln(t)} \right\} \right)
\]

\[
\leq \sum_{t=0}^{T-1} \sum_{x=\bar{a}_T}^{t} \sum_{y=1}^{t} \left( e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\frac{\Delta_a}{2}\right)^2} \left(\frac{2}{y} \ln(t)\right)^2 \right)
\]

\[
\leq \sum_{t=0}^{T-1} \sum_{x=\bar{a}_T}^{t} \sum_{y=1}^{t} \left( e^{-4 \ln(t)} + e^{-4 \ln(t)} \right) \leq \sum_{t=0}^{T-1} t^2 \left(\frac{2}{t^4}\right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.
\]

- Video lectures, solved quizzes and exams, programming assignments all linked from previous years’ course pages.

  [https://www.cse.iitb.ac.in/~shivaram/teaching/old/cs747-a2021/index.html](https://www.cse.iitb.ac.in/~shivaram/teaching/old/cs747-a2021/index.html)