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
CS 747: Foundations of Intelligent and Learning Agents

Agent

Environment/Task

Algorithm for . . .

Stock trading

AGENT

Think

Sense

Act

ENVIRONMENT

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

- Agent Think
- Environment Sense
- Act

<table>
<thead>
<tr>
<th>Agent</th>
<th>Environment/Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm for . . .</td>
<td>Stock trading</td>
</tr>
<tr>
<td>Autopilot program</td>
<td>Airplane</td>
</tr>
</tbody>
</table>

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

Agent

Algorithm for . . .

Stock trading

Autopilot program

Airplane

You

Bicycle

ENVIRONMENT

AGENT

Think

Sense → Act

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

Agent
Algorithm for . . .
Autopilot program
You
You

Environment/Task
Stock trading
Airplane
Bicycle
Academic programme at IIT Bombay
<table>
<thead>
<tr>
<th>Agent</th>
<th>Environment/Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm for ...</td>
<td>Stock trading</td>
</tr>
<tr>
<td>Autopilot program</td>
<td></td>
</tr>
<tr>
<td>You</td>
<td></td>
</tr>
<tr>
<td>You</td>
<td></td>
</tr>
<tr>
<td>AlphaGo</td>
<td></td>
</tr>
</tbody>
</table>
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
CS 747: Foundations of Intelligent and Learning Agents

- **Prerequisites**: Probability, Algorithms, Programming.

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)
**Prerequisites:** Probability, Algorithms, Programming.

Coming up in two weeks!

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

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

\[ \leq \sum_{t=0}^{T-1} \sum_{x=\overline{u}_a^T} t \sum_{y=1}^T \left( e^{-2x} \left( \frac{\Delta_a}{2} \right)^2 + e^{-2y} \left( \sqrt{\frac{2}{y} \ln(t)} \right)^2 \right) \]

\[ \leq \sum_{t=0}^{T-1} \sum_{x=\overline{u}_a^T} t \sum_{y=1}^T \left( e^{-4 \ln(t)} + e^{-4 \ln(t)} \right) \leq \sum_{t=0}^{T-1} t^2 \left( \frac{2}{t^4} \right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}. \]
Prerequisites: Probability, Algorithms, Programming.

Coming up in two weeks:

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

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

\[
\leq \sum_{t=0}^{T-1} \sum_{x=\tilde{u}^*_a} \sum_{y=1}^t \left( e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y} \ln(t)}\right)^2} \right)
\]

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

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