高级算法
Advanced Topics in Algorithms

陈 旭

数据科学与计算机学院

中山大學

SUN YAT-SEN UNIVERSITY

Stochastic Games

- Under the stochastic game, there is a set of states $\boldsymbol{X}$ with a stage game defined in each state.
- In each state x, player i chooses actions from a set $\boldsymbol{A}_i(x)$.
- One of these stage games is played at each of the discrete time $t = 0, 1, 2, \ldots$.
- Informally,
  - Given the system in state $x \in \boldsymbol{X}$. players choose actions $a_1 \in \boldsymbol{A}_1(x)$ and $a_2 \in \boldsymbol{A}_2(x)$.
  - Player i receives a reward of $r_i(x, a_i, a_{-i})$.
  - The probability that they find state x' in next discrete time is $p(x'|x, a_1, a_2, \ldots, a_n)$.

Stochastic Games

Stochastic games generalize both Markov decision processes (MDPs) and repeated games：

- An MDP is a stochastic game with only 1 player
- A repeated game is a stochastic game with only 1 state

Definition

A strategy is called a Markov strategy if the behavior of a player at time t depends only on the state x. A pure Markov strategy specifies an action $a(x)$ for each state $x \in \boldsymbol{X}$.

Assumptions

To simplify discussion, we make the following assumptions:

- The length of the game is not know to the players (i.e., infinite horizon).
- The rewards and transitions are time-independents.
- The strategies of interests are Markov.

Payoff

- Consider a game in state x at time t.
- If we know the NE strategies for both players from $t+1$ onwards, we could calculate the expected future payoffs given that they start from state x. Let $\pi_i^*(x)$ be the expected future payoff for player i starting in state x. (note: the * indicates that these payoffs are derived using the NE strategies for both players).
- At time t, both players would then be playing a single-decision game with payoffs:

$$\pi_i(a_1, a_2) = \left(r_1(x, a_1, a_2) + \delta \sum_{x' \in \boldsymbol{X}} p(x' | x, a_1, a_2) \pi_i^*(x') \right) .$$

Payoff: continue

- The payoffs for a Markov-strategy Nash equilibrium are given by the joint solutions of the following pair of equations (one for each state $x \in \boldsymbol{X}$):

$$
\begin{aligned}
\pi_1^*(x) &= \max_{a_1 \in \boldsymbol{A}_1(x)} \left(r_1(x, a_1, a_2^*) + \delta \sum_{x' \in \boldsymbol{X}} p(x'|x, a_1, a_2^*)\pi_1^*(x') \right), \\
\pi_2^*(x) &= \max_{a_2 \in \boldsymbol{A}_2(x)} \left(r_2(x, a_1^*, a_2) + \delta \sum_{x' \in \boldsymbol{X}} p(x'|x, a_1^*, a_2)\pi_2^*(x') \right).
\end{aligned}
$$

- In general, solving these equations can be computationally expensive !!