Teoria dos Jogos

Hokama PhD

11 de maio de 2023

Referências

- Game Theory, Stanford University on Coursera. Matthew O. Jackson, Yoav Shoham e Kevin Leyton-Brown. https://www.coursera.org/learn/game-theory-1
- Algorithmic Game Theory, Stanford Fall 2013, Tim Roughgarden. https://timroughgarden.org/f13/f13.html
- Twenty lectures on algorithmic game theory. Tim Roughgarden. Cambridge University Press, 2016.
- Tópicos da Teoria dos Jogos em Computação. In: Anais do 30 o Colóquio Brasileiro de Matemática. Schouery, R. C. S., Lee, O., Miyazawa, F. K., e Xavier, E. C. Rio de Janeiro. Editora do IMPA, 2015.

Jogo Fictício - Convergência

Teorema

Se a distribuição empírica da estratégia de cada jogador converge em um jogo fictício, então converge para um equilíbrio de Nash.

Teorema

Cada uma das seguintes condições é suficiente para a convergência:

- O jogo é de soma zero;
- O jogo é solucionável por dominância;
- O jogo é um jogo potencial;
- O jogo é $2 \times n$ e tem recompensas genéricas.

Aprendizagem sem arrependimento

- Nessa abordagem não observamos diretamente o comportamento dos adversários.
- Começamos modelando um critério que deve ser satisfeito. Chamado de critério sem arrependimento.
- Vamos modelar o arrependimento de um jogador no tempo t como a diferença da utilidade que ele ganhou u^t e da utilidade $u^t(s)$ que ele poderia ter ganhado jogando s.

Definição: Arrependimento

O **arrependimento** que um agente experiência no tempo t por não ter jogado a estratégia s é

$$R^t(s) = u^t(s) - u^t.$$

Def.: Regra do aprend. sem arrependimento

Uma regra de aprendizado apresenta **não arrependimento** se para qualquer estratégia pura s de um agente, vale que

$$Pr\left(\left[\lim_{t\to\infty} R^t(s)\right] \leq 0\right) = 1.$$

Equilíbrio em Jogos repetidos

Estratégias

- O que é uma estratégia pura em um jogo repetido?
  - Uma escolha de ação em todo ponto de decisão.
  - Ou seja, uma ação em todo estágio do jogo.
  - O que é um número infinito de ações.
  - E ainda, em cada escolha, o jogador conhece todas as jogadas anteriores dele e dos adversários, então existe uma escolhe em cada possível história.
- Considere um jogo como o Dilema do Prisioneiro repetido. Estratégias famosas:
  - **Olho-por-olho** (Tit-for-tat): Começa cooperando. Se o adversário delatar, você delata no próximo estágio. Depois volta a cooperar.
  - **Gatilho** (Trigger): Começa cooperando, se o adversário delatar, delata pra sempre.

Regret Matching

- Exemplo de uma regra de aprendizado que aprensenta não arrependimento: **Regret Matching**.
- A cada passo, cada ação é escolhida com probabilidade proporcional ao seu arrependimento. Ou seja,

$$\sigma_i^{t+1}(s) = \frac{R^t(s)}{\sum_{s' \in S_i} R^t(s')}$$

em que $\sigma_i^{t+1}(s)$ é a probabilidade que o agente i jogue a estratégia pura s no tempo $t+1$.
- Converge para um equilíbrio correlacionado para jogos finitos repetidos.

Equil. de Nash em Jogos Repetidos

- Mostramos anteriormente que qualquer jogo finito tem um equilíbrio de Nash (estra. mista).
- Entretanto de tentarmos escrever o jogo repetido em sua forma normal, temos uma quantidade infinita de estratégias puras.
- Portanto o jogo deixa de ser finito, e o teorema de Nash não mais se aplica. Então será que temos um equilíbrio?
- Por outro lado, como temos um número infinito de estratégias puras, será que podemos ter um número infinito de equilíbrios de estratégia pura?

- Nós podemos caracterizar um conjunto de **recompensas** que são alcançáveis no equilíbrio, sem ter que enumerar os equilíbrios.

Exemplo: Considere o seguinte jogo.

	c	d
a	2, 0	0, 0
b	0, 0	0, 2

O perfil de recompensa $(-1, -1)$ é aceitável? Não, pois $v_1 = 0$ e $v_2 = 0$.
O perfil de recompensa $(1, 1)$ é viável? Sim, com $\alpha_{\langle a,c \rangle} = .5$ e $\alpha_{\langle b,d \rangle} = .5$.
O perfil de recompensa $(2, 2)$ é viável? Não

- Considere um jogo de n jogadores $G = (N, A, u)$ e um vetor de recompensas $r = (r_1, r_2, \ldots, r_n)$.
- $v_i = \min_{s_{-i} \in S_{-i}} \max_{s_i \in S_i} u_i(s_{-i}, s_i)$, é o valor minmax do jogador i.

Definição

Um perfil de recompensa r é **aceitável** se $r_i \geq v_i$.

Definição

Um perfil de recompensa r é **viável** se existe valores $\alpha_a \in \mathbb{Q}$ e $\alpha_a \geq 0$ tal que para todo i, podemos expressar r_i como $\sum_{a \in A} \alpha_a u_i(a)$, com $\sum_{a \in A} \alpha_a = 1$.

Exercício

	Cinema	Casa
Cinema	3,0	1,2
Casa	2,1	0,3

No jogo acima, quais recompensas não são aceitáveis?

a (0,3)
b (3,0)
c (2,1)
d (3,3)

Teorema Popular

Teorema Popular (Folk Theorem)

Considere um jogo G com n-jogadores e um vetor de recompensas $(r_1, r_2, \ldots, r_n)$

1. Se r é a recompensa em qualquer equilíbrio de Nash de um jogo repetido com recompensas médias, então para cada jogador i, $r_i \geq v_i$, e portanto r é aceitável.
2. Se r é viável e aceitável, então r é a recompensa de algum equilíbrio de Nash de um jogo repetido G com recompensas médias.

Recompensa em um equilíbrio $\rightarrow$ aceitável

Suponha por contradição que r não é aceitável, isto é, para algum i, $r_i < v_i$.Então considere um desvio desse jogador i para $b_i(s_{-i}(h))$ para qualquer história h do jogo repetido, em que b_i é qualquer melhor resposta no estágio do jogo e $s_{-i}(h)$ é a estratégia dos outros jogadores dado a história atual h.Pela definição de uma estratégia minmax, o jogador i vai receber uma recompensa de pelo menos v_i em todo estágio do jogo se ele adotar essa estratégia e então a recompensa média de i também é pelo menos v_i. Dessa forma i não pode receber a recompensa $r_i < v_i$ em um equilíbrio de Nash.

Viável e aceitável $\rightarrow$ Equilíbrio de Nash

Como r é viável, e os valores α são racionais, nós podemos escreve-los como $r_i = \sum_{a \in A} \left(\frac{\beta_a}{\gamma}\right) u_i(a)$, em que β_a e γ são inteiros não negativos e $\gamma = \sum_{a \in A} \beta_a$.
Iremos construir um perfil de estratégia que vai *ciclar* por todos os resultados $a \in A$ de G com ciclos de comprimento γ, cada ciclo repetindo a ação a exatamente β_a vezes. Seja (a^t) essa sequencia de resultados.

Exemplo, na tabela abaixo estão escritos os $\left(\frac{\beta_a}{\gamma}\right)$

	d	f	g
a	$\frac{2}{7}$	0	0
b	0	$\frac{1}{7}$	$\frac{2}{7}$
c	0	$\frac{2}{7}$	0

Idealmente $J1$ e $J2$ jogariam:
$s_1' = \{a, a, b, b, b, c, c, a, a, b, b, b, \ldots\}$
$s_2' = \{d, d, f, g, g, f, f, d, d, f, g, g, \ldots\}$

Vamos definir a estratégia s_i do jogador i para ser a versão gatilho de jogar (a^t): Se ninguém desviar, então s_i joga a_i^t no período t.
Contudo, se existir um período t' em que algum jogador $j \neq i$ desviou, então s_i vai jogar $(p_{-j})_i$, em que (p_{-j}) é a solução para o problema de minimização na definição de v_j

Primeiro observe que se todo mundo jogar de acordo com s_i, então, por construção, o jogador i recebe a recompensa média de r_i (observe as médias nos períodos de comprimento γ).
Depois, esse perfil de estratégia é um equilíbrio de Nash. Suponha que todo mundo jogue de acordo com s_i, e o jogador j desvia em algum ponto. Então, depois e para sempre, o jogador j vai receber o seu minmax $v_j \leq r_j$, fazendo o desvio não lucrativo.

Jogos Repetidos Descontados

- A anedota é que na década de 1950 todo mundo conhecia o Folk Theorem, mas ninguém sabia apontar o autor.
- Não é necessário que r_i seja racional, mas isso deixa a matemática mais simples.

- O futuro é incerto, estamos frequentemente motivados pelo que acontece hoje.
- O balanço entre o hoje e o futuro é importante em como vamos nos comportar hoje.
- Haverá punição caso alguém se comporte mal?
  - Os outros tem interesse nisso?
  - O infrator se importa?

- Considere um jogo $G = (N, A, u)$ que será jogado repetidamente
- Fatores de desconto $\beta_1, \ldots, \beta_n$, $\beta_i \in [0, 1]$
- Por agora considere os fatores todos iguais: $\beta_i = \beta$ para todo i.
- Recompensa de jogadas $a^0, a^1, a^2, \ldots$ para o jogador i:

$$\sum_t \beta_i^t u_i(a^t)$$

- Histórias de comprimento t: $H^t = \{h^t : h^t = (a^1, \ldots, a^t) \in A^t\}$
- Todas as histórias finitas: $H = \bigcup_t H^t$
- Uma estratégia: $s_i : H \to \Delta(A_i)$, em que $\Delta(A_i)$ são todas as estratégias (de 1 estágio), puras ou mistas.

Dilema do Prisioneiro

Em Jogo Repetido

- $A_i = \{C, D\}$
- Uma história de três períodos: (C, C), (C, D), (D, D).
- Uma estratégia para o 4 período precisaria especificar o que o jogador faz depois de ver a história (C, C), (C, D), (D, D) jogada nos três primeiros períodos.

Exercício

- Em um jogo de Pedra-Papel-Tesoura. Quantos elementos tempos em H^2?
- Resposta: 9^2

Perfeição em Subjogo

Em Jogo Repetido

- Perfis de estratégias que são equilíbrios de Nash em todo subjogo
- Um Subjogo é uma calda do jogo, ou seja, a partir de qualquer período até o limite no infinito.
- Repetidamente jogar um equilíbrio de Nash do jogo de 1 estágio é sempre um equilíbrio perfeito em subjogo para o jogo repetido.

- Em um caso mais interessante em que β não é zero. Qual a recompensa de um jogador casos todos cooperem para sempre, e se ele delata?

$$\begin{array}{lll} \text{Coopera:} & 3 + & \beta 3 + \beta^2 3 + \beta^3 3 + \ldots = \frac{3}{1-\beta} \\ \text{Delata:} & 5 + & \beta 1 + \beta^2 1 + \beta^3 1 + \ldots = 5 + \beta\frac{1}{1-\beta} \\ \hline \text{Diferença:} & -2 + & \beta 2 + \beta^2 2 + \beta^3 2 + \ldots = \beta\frac{2}{1-\beta} - 2 \end{array}$$

- Diferença é não negativa se $\beta\frac{2}{1-\beta} - 2 \geq 0$, ou seja, $\beta \leq \frac{1}{2}$
- Se o jogador se importa com o amanhã pelo menos metade do que se importa com o hoje, não vale a pena a delação.

Considere o seguinte Dilema do Prisioneiro

	C	D
C	3,3	0,5
D	5,0	1,1

- Cooperam desde que todos tenham cooperado no passado.
- O jogadores delatam para sempre depois que alguem delata: **Gatilho sombrio** (Grim Trigger)
- Se $\beta = 0$ qual o equilíbrio do jogo?

E se tornarmos a traição mais atraente:

	C	D
C	3,3	0,10
D	10,0	1,1

$$\begin{array}{lll} \text{Coopera:} & 3 + & \beta 3 + \beta^2 3 + \beta^3 3 + \ldots = \frac{3}{1-\beta} \\ \text{Delata:} & 10 + & \beta 1 + \beta^2 1 + \beta^3 1 + \ldots = 10 + \beta\frac{1}{1-\beta} \\ \hline \text{Diferença:} & -7 + & \beta 2 + \beta^2 2 + \beta^3 2 + \ldots = \beta\frac{2}{1-\beta} - 7 \end{array}$$

- Diferença é não negativa se $\beta\frac{2}{1-\beta} - 7 \geq 0$, ou seja, $\beta \leq \frac{7}{9}$
- Se o jogador se importa com o amanhã pelo menos $\frac{7}{9}$ do que se importa com o hoje, não vale a pena a delação.

- Jogar uma estratégia com recompensas relativamente altas.
- Se alguém desviar, punir esse jogador com uma estratégia que tem recompensas negativas para esse jogador (O que incentiva ele a cooperar)
- Não pode ser uma ameaça vazia, deve ser um equilíbrio no subjogo restante.