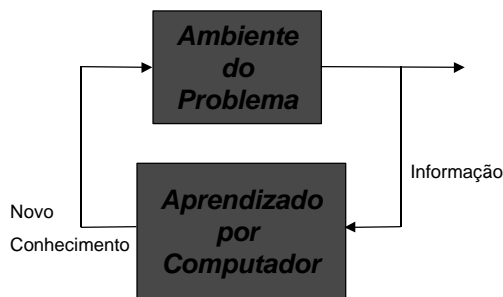


# Computação Evolucionária em Machine Learning

Programas capazes de construir novo conhecimento ou de aperfeiçoar conhecimento existente, usando informação de entrada.



•Aplicações:

- Jogos
- Robótica
- Biologia e Medicina
- Engenharia
- Ciências Sociais



## Dilema dos Prisioneiros

- Algoritmo Genético é usado para “aprender” uma estratégia para um jogo.
- **INDIVÍDUO** ≡ estratégia de jogo, regra de comportamento
- **APTIDÃO** ≡ função da interação com outros jogadores (pontuação)
- **AMBIENTE** ≡ interativo entre indivíduos coevolventes
- Problema é usado para estudar fatores associados com a evolução de cooperação e agressão em comunidades sociais. (Merrill Flood & Melvin Dresher 1950s)



## Dilema dos Prisioneiros

- Dois suspeitos de terem cometido um crime estão em celas separadas e a polícia propõe um acordo. O que pode acontecer?
- Os prisioneiros têm duas opções:
  - **Delatar** : fazer um acordo com a polícia e delatar o parceiro
  - **Cooperar** : manter silêncio sobre o delito e cooperar com o parceiro
- O que pode acontecer ?
  - Nenhum aceita trair
  - Apenas um trai
  - Ambos traem



## Recompensa = Máximo - Penalidade

- Nenhum aceita trair:
  - ambos cooperam e recebem pequena punição (2 anos) por falta de provas; **Recompensa intermediária**=  $(5-2) = 3$
- Apenas um trai:
  - o traidor é libertado (0 anos); o outro é punido (5 anos) ; recompensa por trair é uma **Tentação**= $(5-0) = 5$ ; recompensa do ingênuo (**Sucker**) baixa=  $(5-5)=0$ .
- Ambos traem:
  - punição intermediária (**Punishment**) para ambos (4 anos) = $(5-4)=1$



## Tabela de Recompensas

		Jogador B		Jogador B	
		C	D	C	D
Jogador A	C	$(\gamma_1, \gamma_1)$	$(\gamma_2, \gamma_3)$	$(3, 3)$	$(0, 5)$
	D	$(\gamma_3, \gamma_2)$	$(\gamma_4, \gamma_4)$	$(5, 0)$	$(1, 1)$

**D** = Delatar  
**C** = Cooperar

**D** = Delatar  
**C** = Cooperar

Restrições:

$\gamma_3 > \gamma_1 > \gamma_4 > \gamma_2$  → delatar é mais atraente do que cooperar porém,

$2\gamma_1 > \gamma_2 + \gamma_3$  → cooperar aumenta a recompensa de ambos a longo prazo

$\gamma_2 + \gamma_3 > 2\gamma_4$  → se ambos sempre delatam o resultado é ainda pior



## Características do DP

- jogo não cooperativo para 2 jogadores
- pode ser disputado em torneio entre vários jogadores
- Axelrod promoveu 2 torneios mundiais de estratégias p/ DP
- Estratégia vencedora: Tit\_for\_Tat (Anatol Rapoport)
  - coopera na primeira jogada e depois repete a atitude do oponente na jogada anterior.
  - “Coopera no primeiro encontro e a seguir retribui na mesma moeda”
- Axelrod usou Algoritmos Genéticos para evoluir novas estratégias; as 8 melhores estratégias (humanas) dos torneios serviram para avaliar os indivíduos (ambiente de avaliação estático).



# Modelagem do GA

- **Indivíduo** (Estratégia)

- um indivíduo do GA representa uma estratégia de um jogador cuja atitude é função dos 3 últimos resultados (história).

- **Representação**

- Ao final de cada jogada podemos ter 4 possibilidades:
  - Os dois jogadores delataram: **DD 11 Punishment**
  - Apenas o jogador A delatou: **DC 10 Temptation**
  - Apenas o jogador B delatou: **CD 01 Sucker**
  - Nenhum jogador delatou: **CC 00 Reward**
- Nas últimas 3 jogadas há:  $4 \times 4 \times 4 = 64$  histórias diferentes
- Cromossoma possui 64 bits: 1 ou 0 (D ou C)
- Cada bit define a atitude do jogador para cada uma das 64 histórias
- Posição do bit identifica a história



# Representação

Posição	0	1	2	.....	63
História	CCCCC	DCCCC	CDCCCC	.....	DDDDDD
Base 2	000000	100000	010000	.....	111111
Base 4	RRR	TRR	SRR	.....	PPP
<b>String</b>	0	1	1	.....	1
<b>Decisão</b>	C	D	D	.....	D
	coopera	delata	delata	.....	delata

- Posição no cromossoma corresponde a uma história.
- Conteúdo de cada posição corresponde à atitude do jogador.
- Símbolos da base 4 correspondem às iniciais da tabela de recompensa (Reward, Temptation, Sucker e Punishment).
  - $R=0, T=1, S=2, P=3$
  - Exemplo:  $(RST)_4 = R \times 4^0 + S \times 4^1 + T \times 4^2 = 000110 = 24$



# Representação

- Para fazer a estratégia funcionar no início do jogo, são adicionados 6 bits correspondentes a 3 partidas hipotéticas.

Posição	0	1	2	.....	63	6 bits
História	CCCCC	DCCCC	CDCCCC	.....	DDDDDD	
Base 2	000000	100000	010000	.....	111111	
Base 4	RRR	TRR	SRR	.....	PPP	
String	0	1	1	.....	1	010000
Decisão	C	D	D	.....	D	
	coopera	delata	delata	.....	delata	$(SRR)_4 = 2$

- Atitude na primeira jogada = D.
- Na 2ª e 3ª jogadas utiliza-se parte dos 6 bits e os resultados reais.



# Modelagem do GA

## • Avaliação

- cada indivíduo (estratégia) da população joga com cada um dos 8 oponentes um torneio de 151 partidas

$$A_i = \sum^m p_{i,j} / m$$

$p_{i,j}$ : pontos do jogador i na partida j  
m: total de partidas contra todos oponentes

## • Operadores Genéticos

- crossover e mutação sobre binários

## • Seleção

- avaliação na média ➔ 1 cruzamento
- avaliação acima da média ➔ 2 cruzamentos
- avaliação abaixo da média ➔ 0 cruzamentos



## Resultados

- Indivíduos evoluíram regras de comportamento a partir da interação com outros indivíduos.
- 40 rodadas de 50 gerações de 20 indivíduos.
- O GA evoluiu estratégias que venceram Tit-for-Tat.
- Cromossomas de aptidão média eram tão bons quanto as melhores heurísticas.
- Características das estratégias:
  - traem no 1º e no 2º movimentos;
  - sabem pedir desculpas e entrar em cooperação;
  - têm comportamento diferenciado para indivíduos traidores e não-traidores



## Padrões encontrados

- Maioria dos indivíduos apresentava os seguintes padrões:
- **C após (CC) (CC) (CC)**
  - *“Não deixei o barco virar, continue cooperando”.*
- **D após (CC) (CC) (CD)**
  - *“Aceite a provocação, traia depois que outro traiu por nada”.*
- **C após (CD) (DC) (CC)**
  - *“Aceite as desculpas, coopere após cooperação ser restabelecida”.*
- **C após (DC) (CC) (CC)**
  - *“Coopere quando cooperação mútua é restabelecida depois de uma agressão”.*
- **D após (DD) (DD) (DD)**
  - *“Aceite a provocação, traia após três agressões”.*



## Segundo GA

- Axelrod desenvolveu um segundo experimento, permitindo que os indivíduos jogassem uns contra os outros e contra si mesmos (ambiente de avaliação dinâmico).
- Nas primeiras gerações, estratégias cooperativas não encontravam reciprocidade e tendiam a desaparecer.
- Após 10 a 20 gerações, o panorama se revertia: GA encontrava estratégias de cooperação recíproca, que puniam traição.
- Essas estratégias não foram derrotadas pelas menos cooperativas e conseguiram proliferar nas gerações seguintes. 