



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

PROGRAMA INSTITUCIONAL DE BOLSAS DE INICIAÇÃO CIENTÍFICA –
PIBIC

Avaliação de estratégias marcovianas no Dilema do Prisioneiro, o Jogo do Ultimátum e do Ditador

Área do conhecimento: Psicologia
Subárea do conhecimento: Psicologia Social
Especialidade do conhecimento: Teoria dos jogos

Relatório Final
Período da bolsa: de 08/2017 a 07/2018

Este projeto é desenvolvido com bolsa de iniciação científica

PIBIC/CNPq

Orientador: Hector Julian Tejada Herrera
Autor: Katharine Padilha de Paulo



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Sumário

1	INTRODUÇÃO	3
1.1	Normas e Metanormas	4
1.2	Dilema do Prisioneiro	6
1.3	Dilema do caçador	7
2	OBJETIVOS	8
3	METODOLOGIA	8
4	RESULTADOS E DISCUSSÕES	10
4.1	Dilema do caçador	10
4.2	Dilema do prisioneiro com normas	10
5	CONCLUSÕES	14
6	PERSPECTIVAS	14
7	REFERÊNCIAS BIBLIOGRÁFICAS	15
	Apêndices	17
A	Dilema do Prisioneiro	17
B	Cadeias de Markov	18
C	Algoritmos Genéticos	21



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

1 INTRODUÇÃO

Dilemas sociais (DS) são situações nas quais jogadores têm ganhos individuais, e que caso eles optem por desertar ao invés de cooperar receberão um prêmio maior. No entanto, do ponto de vista do grupo, todos estariam melhores se cooperassem ao invés de desertar. [8] O dilema do prisioneiro, o jogo do ultimátum, o jogo do ditador, o jogo do caçador, a tragédia dos comuns e a tragédia do bem público, são dilemas que vêm sendo estudados em áreas como psicologia ao longo dos anos, e modelam situações sociais que permitem o estudo do comportamento social.[8]

Desde a década de 70, a teoria dos jogos tem sido aplicada no estudo do comportamento das espécies [1] por meio do dilema do prisioneiro, e este dilema tem se mostrado uma ferramenta chave para o estudo das interações sociais. Hoje, aplicado em diversas áreas do conhecimento, como a psicologia, filosofia, ciências da computação, economia e biologia. [2]

Alguns DS podem ser modelados como processos estocásticos de memória curta, mais especificamente usando Cadeias de Markov, entre eles, o Dilema do Prisioneiro. Nessa modelagem matemática, a tomada de decisões depende apenas do último estado, última interação, e não da sequência de acontecimentos anteriores. Dessa forma, em cada nova rodada do jogo, cada participante tomará uma decisão se baseando na última rodada. A partir disso, é possível combinar uma abordagem matemática à uma abordagem computacional, modelando esses dilemas sociais e reproduzir situações sociais de múltiplos indivíduos permitindo sua observação e estudo. [3]

É no contexto apresentado acima que surge este projeto, como continuação dos projetos **PVD3766-2015** e **PVD4113-2016** que buscaram modelar e estudar dilemas sociais como modelos computacionais. Permitindo a avaliação de diferentes estratégias, como elas evoluem, os contextos nos quais elas surgiram e em especial, como essas estratégias nos ajudam entender o comportamento social. Os projetos anteriores tiveram como foco o Dilema do Prisioneiro, neles foi possível descobrir estratégias ótimas do ponto de vista evolutivo, além de entender como as estratégias surgem em ambientes controlados.

Neste projeto, continuamos com a mesma abordagem, adicionado elementos que tornam mais complexa a interação social. Entre eles, o custo que o grupo tem que pagar quando um dos seus membros deserta, e a observação desse comportamento por parte de um outro membro quem decidirá repreender ou não o desertor. Este comportamento pode ser comparado à pessoas que jogam lixo na rua, e alguém que observa e repreende ou deixa de repreender.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Esses novos elementos, a perda do grupo e a punição por do desertor configuram o chamado sistema de normas, que quando implantado num grupo, regula as interações sociais [19]. Não entanto, como acontece frequentemente, nem todas as pessoas punem ao infrator de uma norma. Além disso, nem sempre uma pessoa vai se sentir à vontade para repreender alguém que jogou lixo na rua, pois a aplicação da norma também tem um preço, como por exemplo um reação rude por parte da pessoa repreendida.

As relações entre o custo de aplicar a norma, o custo da punição e o custo que o grupo tem que pagar pela deserção, regularam os efeitos da norma. Nesse contexto, uma alternativa de regulação é introduzir um novo nível de observador quem punirá aquele que, tendo observado o desertor e não fez nada. Essa interação é chamada de Metanorma, e ela seguiria os mesmos parâmetros da norma: um custo para sua aplicação e um custo de punição.

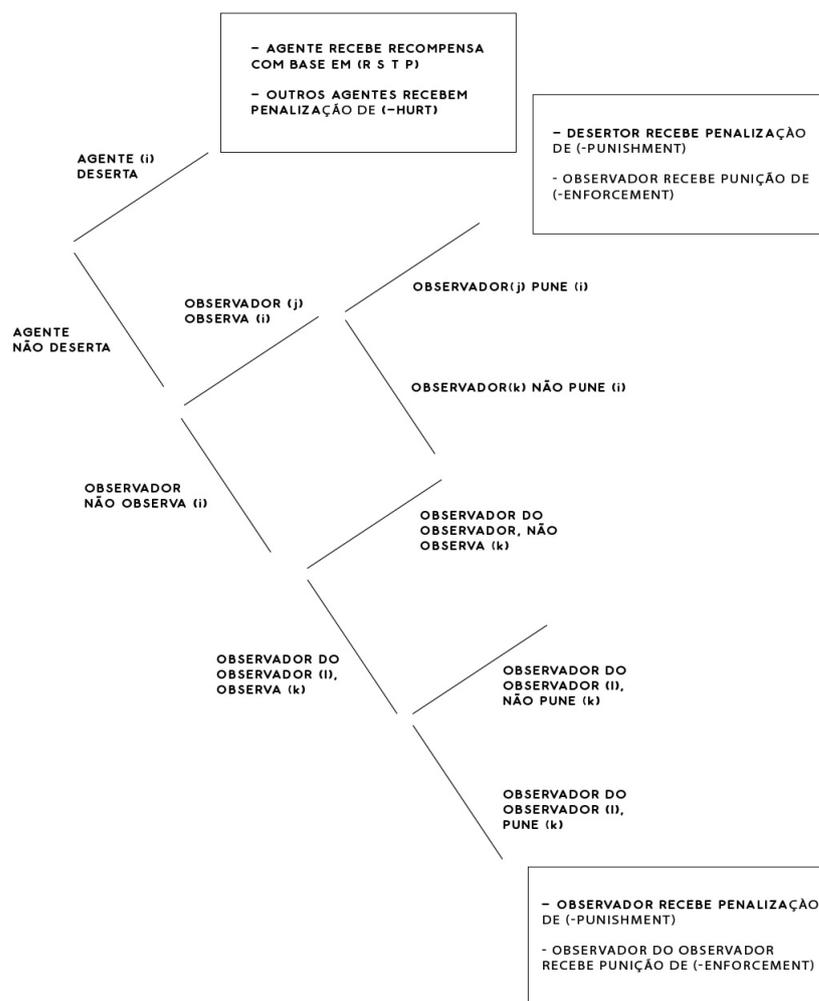
A implantação de sistemas de normas e metanormas deverá, em princípio, garantir o surgimento de estratégias mais cooperadoras, sendo esse um dos principais objetivos do presente projeto. A continuação serão apresentados tópicos nos quais se aprofundará nos modelos teóricos que serão usados e na sua implementação matemática/computacional.

1.1 Normas e Metanormas

Normas contribuem como um modo de regular conflitos em grupos, elas consistem no sentimento de obrigação que os indivíduos têm de seguir algum comportamento, devido ao medo de punições não legais. [19] Elas contribuem em grande parte da vida política e social dos seres humanas, exemplo destas contribuições estão desde o ato de não "colar" em exames, à proibição de armas químicas e o retardo dos programas nucleares. A metanorma, tem o intuito de punir aqueles que decidiram não punir os desertores. Ou seja, aqueles que não aplicaram a norma, ou com palavras populares, "fizeram vista grossa" diante da situação errada. [20]



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA



(a) Esquema de implementação de normas e metanormas no DPI

A política de normas e metanormas serão implementada segundo a figura 1a, um jogador pode desertar ou não, aqueles que não desertem poderão ser observadores que decidirão entre punir ou não desertores que estão ao seu redor. Existe uma probabilidade de desertar e ser observado. Além disso, os jogadores que desertaram serão punidos, e aqueles que escolherem punir os jogadores também terão uma punição por exercer o ato de punir.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Além disso, aqueles jogadores que não são desertores e não foram sorteados como observadores, serão os aplicadores da metanorma, punindo observadores omissos. Assim como a norma, existe uma probabilidade dos agentes que exercem a metanorma de punir ou não os observadores. E os agentes da metanorma também sofreram uma punição por exercer o poder de punir.

Assim, temos as seguintes punições, ou custos por exercer a norma e a metanorma. O custo por exercer a norma (EC), o custo por desertar e ser observado (P), o custo que todos os agentes sofrem quando um agente deserta na rodada (H), o custo por exercer a metanorma (EC') e o custo por não ter punido um desertor e ser observado (P').

Com a implementação de normas e metanormas no dilema do prisioneiro, esperamos encontrar diferentes estratégias, em especial o surgimento de estratégias cooperadoras. Além disso, observar o comportamento e emergência das estratégias ao alterar as probabilidades e os valores das punições aplicadas diante das ações tomadas pelos agentes.

1.2 Dilema do Prisioneiro

O Dilema do prisioneiro (DP) ilustra o interrogatório de dois possíveis prisioneiros. Neste dilema dois suspeitos são presos pela polícia e, enquanto interrogados separadamente, surge a opção de trair o companheiro ou permanecer em silêncio. [7] Caso o suspeito A e o suspeito B decidam permanecer em silêncio ambos ficarão apenas 6 meses na cadeia, e se ambos delatarem um ao outro, cada um receberá 5 anos de prisão. No entanto, se um delatar e o outro permanecer em silêncio, quem delatou sai livre e o outro ficará 10 anos encarcerado.

Como tomar a decisão que acarretará o melhor resultado? Com base no dilema, e no sistema de recompensas citados é possível desenvolver estratégias e obter um melhor desempenho no jogo. Uma explicação completa do dilema e da abordagem utilizada neste projeto está apresentado no Apêndice X, retirado do relatório final do PIBIC PVD4113-2016.

Para avaliar a influência das normas e as metanormas no jogo do DP foram feitas simulações do DP Iterado Multi-agente num modelo evolutivo de aproximadamente 1500 gerações após as quais foram identificadas as 10 melhores estratégias. Essas simulações foram repetidas manipulando as variáveis que controlam a aplicação das normas e metanormas: o custo da punição (EC) variando de 0 até 8, o valor da punição (P) variando de 1 até 9, e o custo da deserção (H), ou quanto todos perdem quando alguém deserta, variando de 0 a 5. Em total foram feitas 384 simulações do modelo evolutivo, totalizando a combinação dos diferentes valores das variáveis que previamente mencionadas (EC, P, H) na presença e ausência da metanorma. A continuação são apresentados os resultados.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

1.3 Dilema do caçador

O dilema do caçador, ou stag hunt, é uma história que se tornou um jogo. Ele discute questões como segurança e cooperação. Ele foi proposto pelo filósofo Jean-Jacques Rousseau na obra *A Discourse on Inequality*. [4]

Durante uma caça, cada caçador deve escolher caçar um veado ou uma lebre. No entanto, as escolhas devem ser tomadas sem o conhecimento da decisão do parceiro de caça. Caso o caçador escolha o veado, é necessária a cooperação do parceiro para que haja uma caça bem sucedida. O caçador deve ponderar, não só o valor de uma lebre e o valor do veado, mas quanto cada caçador receberá por sua participação na caça do veado, e qual a probabilidade de caça ser bem sucedida, dada a quantidade de caçadores empenhados.

Ao contrário de outros dilemas como o Dilema do Prisioneiro, a decisão dos agentes são direcionadas pelo benefício mútuo, e não pelo risco. Dessa forma, um caçador que decide caçar um veado tem o risco de não ser bem sucedido devido à não cooperação do outro caçador. No entanto, o caçador que decide caçar a lebre não possui este risco, a recompensa dele não depende da decisão do outro caçador. Porém, o caçador estará abrindo mão de uma possível caça à um veado bem sucedida. Assim, caçar a lebre é dito como equilíbrio risco-dominante. [4]

A cooperação neste jogo depende da confiança. A cooperação em uma caça com vários caçadores é mais difícil que um caça com duas pessoas. [4] Por meio das matrizes de recompensa abaixo é possível determinar que existem duas estratégias puras de Equilíbrio de Nash, uma que é arriscada-dominante, quando ambos decidem caçar a lebre, e a que é recompensa-dominante que acontece quando ambos decidem caçar o veado.

Assim como no dilema do prisioneiro, é possível analisar essa situação, também, de uma maneira mais abstrata, como um jogo de duas opções, cooperar (C) ou desertar (D) (Ver apêndice A).

	Veado	Lebre
Veado	a,a	c,b
Lebre	b,c	d,d

Tabela 1: Matrix genérica do dilema do caçador

	Veado	Lebre
Veado	2,2	0,1
Lebre	1,0	1,1

Tabela 2: Matrix válida do dilema do caçador



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Neste jogo é necessário que as recompensas mantenham este padrão: $a > b \geq d > c$ para que as características se mantenham. Desta forma, a matriz de recompensas do dilema do caçador se torna $(R, S, T, P) = (CC, CD, DC, DD) = (2, 0, 1, 1)$.

Neste dilema é possível visualizar que incentivos egoístas podem induzir a cooperação, uma vez que o veado é um prêmio maior é possível que ambos os jogadores cooperem para obtê-lo. No entanto, em uma caça com mais que dois indivíduos, pode não ser um incentivo suficiente já que a partilha seria uma recompensa talvez menor que uma lebre. Além disso, diante da escolha de um caçador de caçar uma lebre, é melhor que o outro caçador escolha o veado para evitar competição na caça da lebre, que se torna maior em uma situação com mais caçadores.

2 OBJETIVOS

O objetivo da nossa pesquisa é continuar o estudo sobre dilemas sociais que vem sendo feito nos últimos dois anos por esta linha de pesquisa. Com isso, espera-se obter estratégias evolutivamente ótimas do dilema do caçador e do dilema do prisioneiro com a presença de normas e metanormas, a fim de estudar o comportamento destes agentes e entender em quais contextos essas estratégias surgiram. Além disso, comparar os resultados obtidos neste projeto com os resultados obtidos nesta linha de pesquisa nos anos anteriores.

3 METODOLOGIA

Neste presente projeto, foram desenvolvidos modelos computacionais do dilema do caçador e do dilema do prisioneiro com a atuação de normas e metanormas no software Netlogo. O Netlogo vem sendo escolhido por essa linha de pesquisa devido ao seu ambiente de simulação multiagentes com uma linguagem fácil e intuitiva, e por ser um software livre. [18]

Ambos os modelos se baseiam em algoritmos genéticos e cadeias de markov (Apêndice B). Estes modelos consistem em uma sequência de interações entre agentes, que terão como base duas decisões, cooperar ou desertar. Nestes modelos, uma população inicial de 100 agentes interagem segundo estratégias, regras que definem a tomada de decisões de cada agente, inicialmente aleatórias. Essas estratégias seguem o modelo markoviano apresentado no apêndice A.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

A cada geração, a população é filtrada com base em uma função fitness, que define o desempenho de cada agente durante as rodadas daquela geração. Os agentes que tiverem a melhor fitness, estarão na próxima geração. Os agentes evoluirão até o surgimento de estratégias dominantes. Este modelo segue o modelo é criado a partir de um algoritmo genético simples (Ver Apendice C) criado pela autora deste mesmo projeto no PIBIC do ano 2016.

A partir do modelo acima (Modelo evolutivo do Dilema do prisioneiro Iterado Multi-agentes), foram criados o modelo do dilema do caçador e o modelo de normas. O modelo do dilema do caçador é semelhante ao DP, porém com um sistema de recompensas diferente. Enquanto as recompensas do DP segue a matriz (3, 0, 5, 1), a do DC tem uma matriz de (2, 0, 1, 1).

Já o modelo de normas, contam com modificações significativas quanto a interação. Nele, é inserido um sistema de punição, ou normas. Durante a sua execução são escolhidos grupos de observadores, e observadores dos observadores. Observadores são aqueles que poderão punir desertores. Os observadores dos observadores são outros agentes que poderão punir os observadores que optaram por não punir os desertores. O critério de punição foi adicionado da seguinte maneira: a probabilidade de um agente punir depende de sua estratégia. Assim, a probabilidade de uma estratégia punir ou não é definido da seguinte maneira.

$$Vetor_{estratégia} = [RSTP] \quad (1)$$

$$P_{punir} = Vetor[0] + Vetor[1] + Vetor[2] + Vetor[3]/4 \quad (2)$$

Deste modo, para a estratégia TFT: [1 0 1 0], a probabilidade de punir é dada por $(1 + 0 + 1 + 0)/4 = 0,5 = 50\%$, estratégias cooperadoras tem a probabilidade de 100% de punir, enquanto estratégias desertoras tem 0% de chance de punir.

Neste modelo, além da interação comum a todos os modelos, a cada rodada, após a interação dos agentes, a norma e a metanorma, caso ativada, são exercidas. Desta forma os grupos de observadores, e observadores dos observadores tomam suas decisões. Após estas decisões as recompensas pessoais de cada agente são recalculadas, a fitness é atualizada e após a sequência de rodadas a geração é encerrada. Os agentes com melhores estratégias evoluem para a próxima geração e assim sucessivamente até o fim do jogo.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

4 RESULTADOS E DISCUSSÕES

4.1 Dilema do caçador

A partir das simulações realizadas com o dilema do caçador, com a configuração de recompensas apresentada na tabela 2, obtemos a estratégia $[1\ 0\ 0\ 0]$ como estratégia dominante. Analisando o perfil da estratégia com base na matriz das recompensas, vemos que os jogadores tendem a ficar na área de caça ao veado ou de caça a lebre, mostrando que são as duas melhores opções. Ou seja, se ambos cooperarem é mais vantajoso caçar um veado, caso o contrário é mais vantajoso desertar, ou seja, que ambos optem por caçar a lebre.

Após a penalização sucessivas das estratégias dominantes vemos o surgimento de diferentes estratégias, entre elas, $[1\ 0\ 1\ 0]$ e $[1\ 0\ 0\ 1]$, duas estratégias do DP. Isso mostra que elas também se aplicam ao dilema do caçador. A lista completa com as melhores estratégias do dilema do caçador e do dilema do prisioneiro pode ser visualizada na Tabela 3.

Posição	Estratégia DC	Estratégia DP
1º	$[1\ 0\ 0\ 0]$	$[1\ 0\ 0\ 0]$
2º	$[1\ 0\ 1\ 0]$ e $[1\ 0\ 0\ 1]$	$[0000]$ e $[0100]$
3º	$[1\ 0\ 1\ 1]$	$[0001]$
4º	$[1\ 1\ 0\ 1]$ e $[1\ 1\ 1\ 1]$	$[1001]$
5º	$[1\ 1\ 0\ 0]$ e $[1\ 1\ 1\ 0]$	$[1010]$
6º	$[0\ 0\ 1\ 1]$ e $[0\ 1\ 0\ 1]$	$[0010]$, $[1011]$, $[0101]$ e $[0110]$.

Tabela 3: Resultados dos algoritmos genéticos para o dilema do caçador e do prisioneiro

Apesar da presença das estratégias TFT e Pavlov no DC, vemos que as demais estratégias ou surgem em ordens diferentes, ou são diferentes. Por exemplo a estratégia $[1\ 0\ 1\ 1]$ nem aparece no DP. Mostrando que a diferença entre as recompensas dos dilemas mudam a forma como as estratégias evoluem.

4.2 Dilema do prisioneiro com normas

Como foi abordado na metodologia, no modelo de normas, a probabilidade de um agente punir está conectada a estratégia do mesmo. Deste modo, estratégias cooperadoras tem uma chance maior de punir, enquanto estratégias desertoras têm uma chance menor, ou nenhuma, de punir.



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

Com isso, chegamos à um modelo mais real e semelhante a sociedade, onde quem é “anti-ético” não deve ter o direito de julgar o comportamento dos demais. Com as normas e metanormas, foi possível observar o surgimento da cooperação no dilema do prisioneiro iterado.

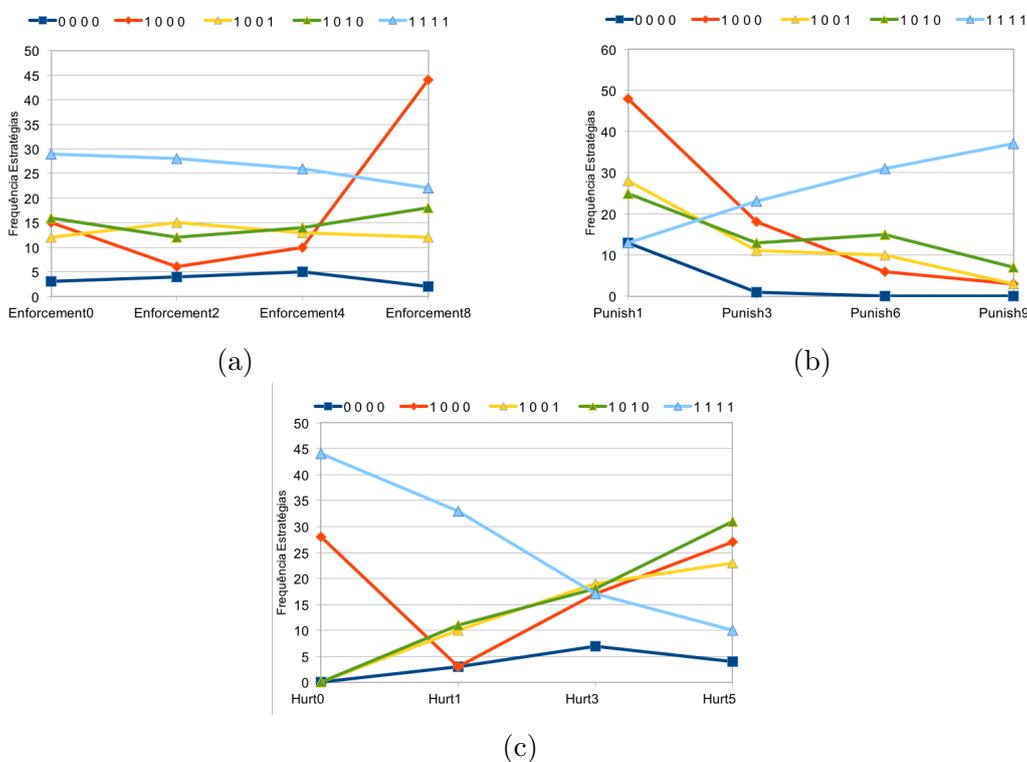


Figura 2: Desempenho de diferentes estratégias no DPI num ambiente controlado por normas medido como a frequência na qual uma determinada estratégia é encontrada dentro do grupo dos dez melhores agentes nas diferentes gerações que foram avaliadas. Em (a), foi manipulado o custo da punição (enforcement) indo de 0 até 8. Em (b), foi manipulado o valor da punição (punish) indo de 1 até 9. E finalmente em (c), foi manipulado o custo da deserção (hurt), ou quanto todos perdem quando alguém deserta, indo de 0 a 5.

Em simulações sem a atuação de metanormas (Figura 2), podemos ver que a medida que o custo do agente por aplicar a norma aumenta (EC) (Figura 2-2a), a estratégia dominante encontrada no modelo do DPI, [1 0 0 0] se torna mais frequente. Isso ocorre pelo seguinte motivo, a probabilidade da estratégia [1 0 0 0] de punir os outros é uma das menores ($1/4 = 25\%$), se comparada às demais estratégias presentes no ambiente, maior apenas que a probabilidade de [0 0 0 0] que é 0%. A estratégia [1 0 0 0] quase não aplica a norma, então não é afetada pela variável (EC).



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

Podemos observar também na figura X-b que, a medida que a punição por desertar aumenta (P), a cooperação se mostra dominante, uma vez que o custo de desertar se torna muito alto. Quando o custo chega a 9, a estratégia [0 0 0 0] não surge nesse ambiente.

Além disso, a medida que aumenta o custo de punição geral (H), as estratégias ficam mais equilibradas, como TFT, [1 0 0 0] e Pavlov predominando. Isso acontece porque ocorre um equilíbrio entre cooperação e deserção. Todas estratégias são penalizadas quando um agente deserta, assim, a medida que essa punição aumenta, todas as estratégias se prejudicam. Além disso, as estratégias desertoras se tornam menos vantajosas porque também sofrem a norma, por exemplo, [0 0 0 0] aparece em último lugar no gráfico 2c. Isso ilustrando situações sociais nas quais pessoas que só pensam em "si mesmo" e que prejudicam os outros não sobrevivem em grupo.

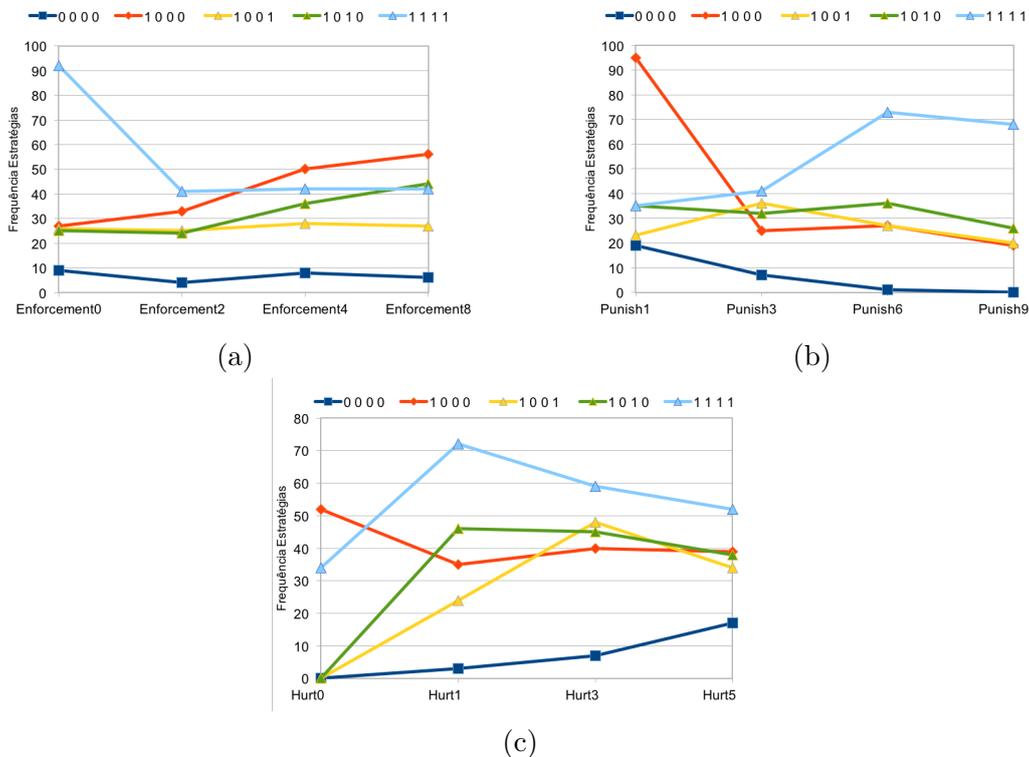


Figura 3: Desempenho de diferentes estratégias no DPI num ambiente controlado por normas e metanormas. Em A, foi manipulado o custo da punição (enforcement) indo de 0 até 8. Em B, foi manipulado o valor da punição (punish) indo de 1 até 9. E finalmente em C, foi manipulado o custo da deserção (hurt), ou quanto todos perdem quando alguém deserta, indo de 0 a 5.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Nas simulações com normas e metanormas (Figura 3) , vemos que a estratégia [1 0 0 0] continua dominante, com as demais estratégias equilibradas logo abaixo. Também verificamos que a estratégia cooperadora [1 1 1 1] continua dominante quando aumentamos o custo da punição (P), enquanto as demais se tornam menos frequentes. Este fato surge com a implementação de metanormas, nas quais agentes com as demais estratégias que optam por não punir os agentes desertores, agora também são punidos. Assim, as outras estratégias se tornam menos vantajosas, enquanto que cooperadores que sempre exercem a norma em desertores, não são penalizados pela metanorma.

No figura 3, vemos que a estratégia [1 1 1 1] evolui significativamente das simulações sem metanormas para o com metanormas. Enquanto as demais estratégia permanecem relativamente iguais. Isso contribui para nossa teoria de que na presença de uma metanorma, a pressão sob os desertores é ainda maior, punindo desertores e aqueles que são condizentes com suas atitudes. Assim, podemos ver que a cooperação se torna a melhor estratégia quando a metanorma está ativada.

Com isso, vimos que, uma estratégia que têm probabilidades maiores, sempre exercerão a norma, de modo a ter seu payoff reduzido por punir desertores (EC), enquanto probabilidades menores não serão penalizadas por exercer a norma. No entanto, estratégias desertoras, por ter a maior chance de desertar serão punidas pela norma por desertar durante o jogo, enquanto as estratégias mais cooperadoras não serão punidas pela norma (P). Além disso, estratégias que podem punir mas não o fazem acabam se prejudicando com a aplicação das metanormas. Assim, a estratégia [1 1 1 1], ou a cooperação, se torna a melhor opção durante o jogo com normas.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

5 CONCLUSÕES

Com este projeto foi possível observar o surgimento de estratégias do DP no dilema do caçador, bem como estratégias nunca vistas no modelo evolutivo do DP, mostrando que apesar de semelhantes, são dilemas com perspectivas diferentes. Além disso, foi possível observar o surgimento da cooperação no DPI, com a utilização de normas e metanormas. Além de analisar o surgimento e a evolução das estratégias ao alterar as variáveis do modelo de normas de Axelrod. Com isso, esperamos continuar implementando variáveis aos modelos, a fim de torná-los cada vez mais próximos de modelos sociais reais e com isso podendo simular e aprofundar os estudos em dilemas e situações sociais.

6 PERSPECTIVAS

Esperamos nos projetos seguintes implementar o modelo de Axelrod em outros dilemas, entre eles o dilema do caçador, que possui uma identidade maior de comunidade que o dilema do prisioneiro. Com isso, observar como as estratégias surgem nos demais dilemas com as normas e metanormas, para tentar entender o comportamento dos agentes no jogo e de pessoas em situações sociais semelhantes. Além disso, esperamos testar as estratégias obtidas em um modelo de torneios para verificar a supremacia das estratégias dominantes obtidas no modelo evolutivo.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

7 REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Matt Ridley. The Origins of Virtue. Penguin Books Limited, 1997.
- [2] Melvin Dresher. Games of strategy. Rand Corporation, 2007.
- [3] Joseph E Harrington. Cooperation in a one-shot prisoners' dilemma. Games and Economic Behavior, 8(2):364–377, 1995.
- [4] Brian Skyrms. The Stag Hunt and the Evolution of Social Structure. Cambridge University Press.
- [5] Alan G Sanfey, James K Rilling, Jessica A Aronson, Leigh E Nystrom, and Jonathan DCohen. The neural basis of economic decision-making in the ultimatum game. Science,300(5626):1755–1758, 2003.
- [6] Uri Wilensky. NetLogo user manual (version 5.3.1), 1999.
- [7] Axelrod, Robert, and William Donald Hamilton. "The evolution of cooperation."science 211.4489 (1981): 1390-1396.
- [8] DAWES, Robyn M.; MESSICK, David M. Social dilemmas. International journal of psychology, v. 35, n. 2, p. 111-116, 2000.
- [9] KOMORITA, S. S.; PARKS, C. D. Social Dilemmas. Westview Press, 1996. 180 p. ISBN 0813330033.
- [10] NOWAK, M.; SIGMUND, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature, Nature Publishing Group, v. 364, n. 6432, p. 56–8, jul 1993. ISSN 0028-0836.
- [11] ROSS, S. M. Introduction to Probability Models. Academic Press, 2006. 800 p. ISBN 0123756871.
- [12] YATES, R. D. Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers, 3rd Edition: Third Edition. John Wiley & Sons, 2014. 544 p. ISBN 1118804384.



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

[13] PRESS, W. H.; DYSON, F. J. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences of the United States of America*, v. 109, n. 26, p. 10409–13, jun 2012. ISSN 1091-6490.

[14] HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. [S.l.]: M.I.T.P., 1992. ISBN 978-0-262-58111-0.

[15] COLEY, D. A. *An Introduction to Genetic Algorithms for Scientists and Engineers*. [S.l.]: World Scientific, 1999. ISBN 978-981-02-3602-1.

[16] MAJESKI, S. J. Arms races as iterated prisoner's dilemma games. *Mathematical Social Sciences*, v. 7, n. 3, p. 253–266, jun. 1984. ISSN 0165-4896

[17] LUCAS, D. C. *Algoritmos genéticos: uma introdução*. 2002.

[18] WILENSKY, U. *NetLogo User Manual (version 5.3.1)*. 1999.

[19] McAdams, R.H., 1997. The origin, development, and regulation of norms. *Mich. Law Rev.* 96 (2), 338–433.

[20] AXELROD, Robert. An evolutionary approach to norms. *American political science review*, v. 80, n. 4, p. 1095-1111, 1986.



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

Apêndices

Apêndices A Dilema do Prisioneiro

O Dilema do prisioneiro (DP) recebe esse nome visto que ilustra o interrogatório de dois possíveis prisioneiros. Neste jogo dois suspeitos são presos pela polícia e, enquanto interrogados separadamente, surge a opção de delatar o companheiro ou permanecer em silêncio. Caso o suspeito A e o suspeito B decidam permanecer em silêncio ambos ficarão apenas 6 meses na cadeia, e se ambos delatarem um ao outro, cada um receberá 5 anos de prisão. No entanto, se um delatar e o outro permanecer em silêncio, quem delatou sai livre e o outro ficará 10 anos encarcerado. Como tomar a decisão que renderá o melhor resultado?

É possível analisar essa situação, também, de uma maneira mais abstrata, como um jogo de duas opções, cooperar (C) ou desertar (D). Cada ação pode gerar diferentes resultados e recompensas (R, S, T, P) que são interpretados da seguinte maneira: *Reward*, quando ambos cooperam (CC); *Punishment*, quando ambos desertam (DD); *Temptation* quando há a tentação de desertar após ambos cooperarem (DC), e *Sucker's Payoff* quando o jogador é traído ao optar por cooperar (CD). Esses resultados podem ser ilustrados por pontos de recompensa. A matriz de recompensas mais comum na literatura é a seguinte $(R, S, T, P) = (CC, CD, DC, DD) = (3, 0, 5, 1)$ [7].

Tabela 4: Recompensas

	<i>Cooperate</i>	<i>Defect</i>
<i>Cooperate</i>	$R = 3, R = 3$	$S = 0, T = 5$
<i>Defect</i>	$T = 5, S = 0$	$P = 1, P = 1$

Ao observar a tabela acima, é possível inferir que a melhor opção pessoal seria desertar, visto que possuem as recompensas mais altas. Porém, diante da quase inevitável deserção mútua, seria melhor coletivamente se ambos cooperassem. Como driblar essa situação e tornar possível a cooperação? Segundo Axelrod [7] a cooperação pode surgir caso o DP seja jogado por um mais de uma vez. Nesse contexto, surge o Dilema do Prisioneiro Iterado. O Dilema do Prisioneiro Iterado (DPI) é uma variação do DP no qual os jogadores se enfrentam inúmeras vezes sob as mesmas circunstâncias. A ação de interagir repetidamente com o mesmo oponente permite o jogador conhecer a estratégia do oponente, possibilitando, deste modo, o surgimento de atitudes como altruísmo e reciprocidade podendo gerar mais tarde a cooperação [9].



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Com base no DPI e a partir do sistema de recompensas citados anteriormente inúmeras estratégias surgiram com o passar dos anos. Entre elas, *Tic-for-tat* (TFT) e *Win-stay Lose-Change*. A primeira coopera ao iniciar o jogo e nas rodadas seguintes repete a decisão anterior do oponente, ela também é conhecida por Olho-por-olho, em tradução livre. Já a segunda *Win-stay Lose-Change*, Vence-fica/perde-muda, pode cooperar ou desertar na primeira jogada, porém nas próximas rodadas ela se comporta da seguinte forma: caso o oponente tenha cooperado na última rodada o jogador mantém a decisão, caso o contrário ele muda. Mantendo os ganhos sempre favoráveis. [10]. Essa última também é conhecida como Pavlov.

Uma similaridade entre ambas estratégias é que elas dependem somente das decisões da última rodada para tomar a decisão seguinte. Essas estratégias se encaixam no modelo Markoviano, no qual o estado seguinte depende unicamente do estado atual. Sob tais condições, é possível escrever essas estratégias como vetores que representam a probabilidade de cooperação do jogador na próxima rodada dado apenas o resultado da rodada anterior.

Além disso, é importante lembrar que para que o DP se mantenha, é necessário que a seguinte expressão $T > R > P > S$ seja verdadeira. E que no DPI, uma segunda expressão, além da citada acima, também deve ser satisfeita $2R > (T + S)$ [9].

Apêndices B Cadeias de Markov

Uma cadeia de Markov é uma sequência discreta de variáveis estocásticas, ou aleatórias.

$$X_1, X_2, X_3 \cdots X_n \quad X_n | n = 0, 1, 2, \cdots \quad (3)$$

Cada X_n denota o estado do processo em um determinado tempo n . Além disso, dado que $X_n = i$, ou seja, que em um determinado tempo n o processo se encontra no estado i , é factível dizer que existe uma probabilidade P_{ij} de que no próximo estado $(n+1)$ o processo esteja em j , $X_{n+1} = j$. [11]

As Cadeias de Markov apresentam uma propriedade característica denominada Memória Markoviana. Ela determina que sendo o estado atual conhecido, apenas ele é relevante para prever o estado seguinte de um processo. Deste modo, o estado atual X_n carrega toda a informação anterior do sistema necessário para prever o próximo estado [12]. Tal propriedade é representada pela seguinte expressão.

$$\begin{aligned} P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \cdots, x_0 = i_0] = \\ P[X_{n+1} = j | X_n = i] = \\ P_{ij} \end{aligned} \quad (4)$$

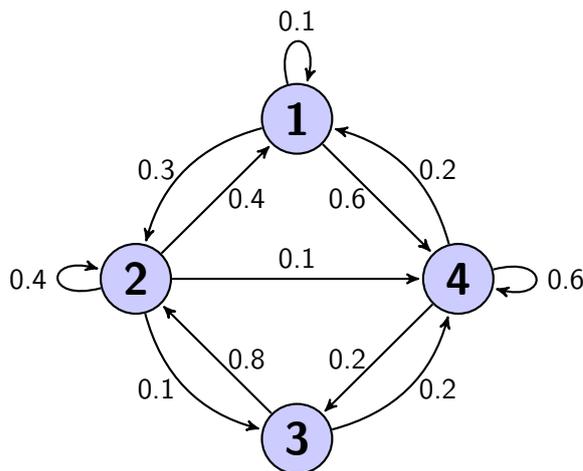


**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Para cada estado existirá uma probabilidade de transição para o estado seguinte e essas probabilidades são agrupadas em uma matriz denominada Matriz de transição. Os eixos (x, y) são definidos por, $P(X_{n+1} = j | X_n = i)$, onde i denota o estado atual e j o próximo estado. [12]

$$P_{i,j} = \begin{pmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,n} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m,1} & P_{m,2} & \cdots & P_{m,n} \end{pmatrix} \quad P_{i,j} \geq 0, \quad \sum_{j=0}^{\infty} P_{i,j} = 1 \quad (5)$$

Para um melhor entendimento é possível, também, representar Cadeias de Markov como um diagrama (Figura 4a). Nele cada nodo representa um estado e cada seta a probabilidade de transição. O diagrama abaixo representa um processo Markoviano de quatro estados (1, 2, 3, 4) e as setas representam a probabilidade de transição de um estado para o outro. Por exemplo, caso o processo esteja no estado 1, a probabilidade de estar em 2 no próximo é de 0.3, enquanto a mudança para o estado 3 é impossível visto que não há setas. Além disso, a matriz de transição ao lado (Figura 4b) do diagrama representa todas as probabilidades representadas pelas setas.



Matriz de Transição

$$P = \begin{pmatrix} 0.1 & 0.3 & 0 & 0.6 \\ 0.4 & 0.4 & 0.1 & 0.1 \\ 0 & 0.8 & 0 & 0.2 \\ 0.2 & 0 & 0.2 & 0.6 \end{pmatrix}$$

(b) Matriz de transição do diagrama ao lado

(a) Diagrama de estados de um processo Markoviano

Figura 4: Exemplo de um processo Markoviano



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Neste contexto, o dilema do prisioneiro iterado pode ser analisado como um processo estocástico Markoviano no qual as estratégias dos jogadores seguirão um mapa de probabilidades condicionais. As estratégias são definidas por probabilidades condicionais p_1, p_2, p_3, p_4 que representam as chances de cooperação na próxima rodada, dado o resultado da última interação dos dois agentes, R, S, T , ou P , respectivamente.

	Y		
		<i>Cooperate</i>	<i>Defect</i>
X			
	<i>Cooperate</i>	R	S
	<i>Defect</i>	T	P

Desse modo, é possível considerar que o dilema do prisioneiro iterado abordado acima é um processo estocástico de quatro estados $xy \in \{CC, CD, DC, DD\}$ [?]. Cada jogador possui uma estratégia condicionada, $p = (p_1, p_2, p_3, p_4)$ para o jogador X e $q = (q_1, q_2, q_3, q_4)$ para o jogador Y . Como cada estratégia tem por base o ponto de vista do jogador, a estratégia X corresponde a $xy \in \{CC, CD, DC, DD\}$ e a estratégia de Y $yx \in \{CC, CD, DC, DD\}$ [13]. É possível reescrever as principais estratégias da literatura como um vetor de probabilidades condicionais.

Estratégia	Probabilidade (R,S,T,P)
<i>Tit – For – Tat(TFT)</i>	$(1, 0, 1, 0)$
<i>Desertar – sempre(ALL – D)</i>	$(0, 0, 0, 0)$
<i>Cooperar – sempre(ALL – C)</i>	$(1, 1, 1, 1)$
<i>Win – stay Lose – Change(PAVLOV)</i>	$(1, 0, 0, 1)$

Como o modelo obedece a um processo Markoviano, é possível determinar a matriz de transição com suas suas dezesseis probabilidades de transição (Figura 5). Ela determinará a probabilidade do próximo estado dado o estado anterior.

$$\begin{matrix}
 & R & S & T & P \\
 \begin{matrix} R \\ S \\ T \\ P \end{matrix} & \begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_2 & p_2(1-q_2) & (1-p_2)q_2 & (1-p_2)(1-q_2) \\ p_3q_3 & p_3(1-q_3) & (1-p_3)q_3 & (1-p_3)(1-q_3) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix}
 \end{matrix}$$

Figura 5: Matriz de transição para o dilema do prisioneiro no modelo Markoviano



**SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DE SERGIPE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA**

Apêndices C Algoritmos Genéticos

Algoritmos genéticos (**AGs**) são algoritmos numéricos, baseados em processos naturais que são utilizados em soluções de problemas de otimização e busca.

Fundamentado principalmente por John Henry Holland em sua obra *Adaptation in Natural and Artificial Systems* [14], os AGs se inspiram em técnicas e princípios da biologia evolutiva, como hereditariedade, seleção natural, recombinação, e mutação genética [15], para resolver problemas de diversas áreas. Basicamente a idéia por trás dos algoritmos genéticos é fazer o que a natureza faz onde envolve seleção natural. [16].

Os AGs iniciam sua busca a partir de uma população aleatória. Cada indivíduo dessa população possui um vetor binário, contendo uma possível solução gerada aleatoriamente, seu código genético. O código genético é uma representação do espaço de busca do problema a ser resolvido. E então o algoritmo usa os operadores de seleção, crossover e mutação para direcionar a população em direção a solução ótima do problema.

A seleção faz o papel da natureza, aplicando pressão, filtrando indivíduos. Os indivíduos que tiverem um desempenho inferior morrem, e aqueles que possuem um desempenho superior são selecionados e passam de geração. Essa ação é desempenhada em geral pela função “fitness”. Dessa forma, o conteúdo genético de cada jogador selecionado passa para os próximos indivíduos por meio de clonagem ou crossover.

O crossover é responsável por permitir a troca de informações, de soluções entre indivíduos, assim como na reprodução em organismos naturais. O crossover pode ser realizado por exemplo pela recombinação em um ponto, em dois pontos, dentre outras formas [17]. Já a clonagem gera novos indivíduos com as mesmas características genéticas dos anteriores. E a mutação surge como uma mudança moderada aleatória em pontos do vector.

Deste modo, como qualquer programa evolutivo, um algoritmo genético precisa conter todos esses requisitos [16]:

- Uma representação genética para possíveis soluções do problema;
- Valores para parâmetros que são utilizados nos AGs, como tamanho da população, probabilidade de aplicação de operadores genéticos;
- Uma forma de criar uma população inicial de potenciais soluções;
- Uma função de avaliação que faz o papel da seleção, ou seja do meio ambiente, denominada: fitness";
- Operadores genéticos que alteram a composição dos indivíduos durante a reprodução;