Projeto de Formatura – Turmas 2019

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

Engenharia de Computação

Tema: **Impacto da cognição na cooperação de agentes de aprendizagem por reforço no dilema do prisioneiro para mais de dois jogadores.**

Introdução

Do ponto de vista biológico cooperação é uma característica vantajosa, espécies com maior cooperação conseguem vencer outras que cooperam menos na competição por recursos como é o caso das formigas argentinas sobre outros insetos e do *Homo Sapiens* sobre o *Neandertal*. Entretanto há dois fatores que dificultam cooperação: medo de ser explorado e se aproveitar da cooperação dos outros. Este dilema foi modelado com *N-Person Prisoners Dilemma* (*NPD*) [1].

Este trabalho simula variantes de agentes que aprendem por reforço, isto é, por tentativa e erro, jogando *NPD* para identificar quais características destes agentes favorecem a cooperação.

Fundamentos

Há duas maneiras principais de variar a cognição de agentes que aprendem por reforço: ajustando a informação que o agente tem disponível ou melhorando a tomada de decisão com base nessa informação [2].

Foram estudados 4 níveis de informação:

- **MemoryLess:** sem informação.
- **MajorTD4:** sabe sua última ação e a ação mais frequente dos oponentes.
- **Selfless:** sabe o número de cooperadores.
- **LevelLearner:** sabe sua última ação e número de cooperadores.

Foram testados 5 diferentes políticas de atuação:

- **ε-greedy:** explora novas ações com probabilidade fixa ε, escolhe melhor ação caso contrário:
- **Linear ε-greedy:** ε decai linearmente.
- **Logarithmic ε-greedy:** ε decai logaritmicamente.
- **Boltzmann:** calcula a probabilidade de se escolher cada ação.
- **Actor-Critic:** aprende a probabilidade de se cooperar em cada situação.

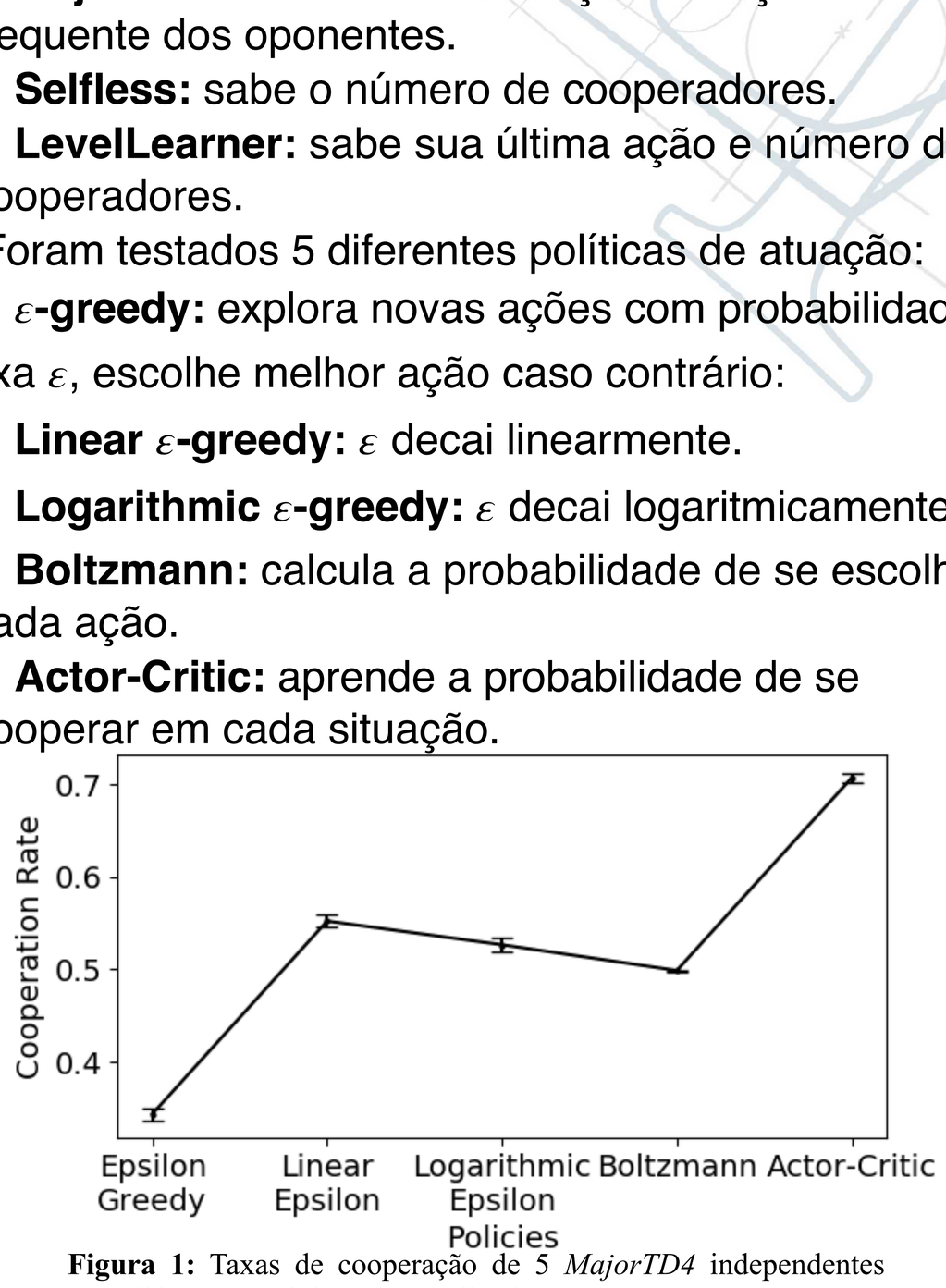

Figura 1: Taxas de cooperação de 5 *MajorTD4* independentes com diferentes políticas de escolha de ações.

Integrantes: João Vitor de Oliveira Barbosa.
Orientador: Anna Helena Reali Costa.
Co-orientador: Francisco C. Santos, Francisco Melo.

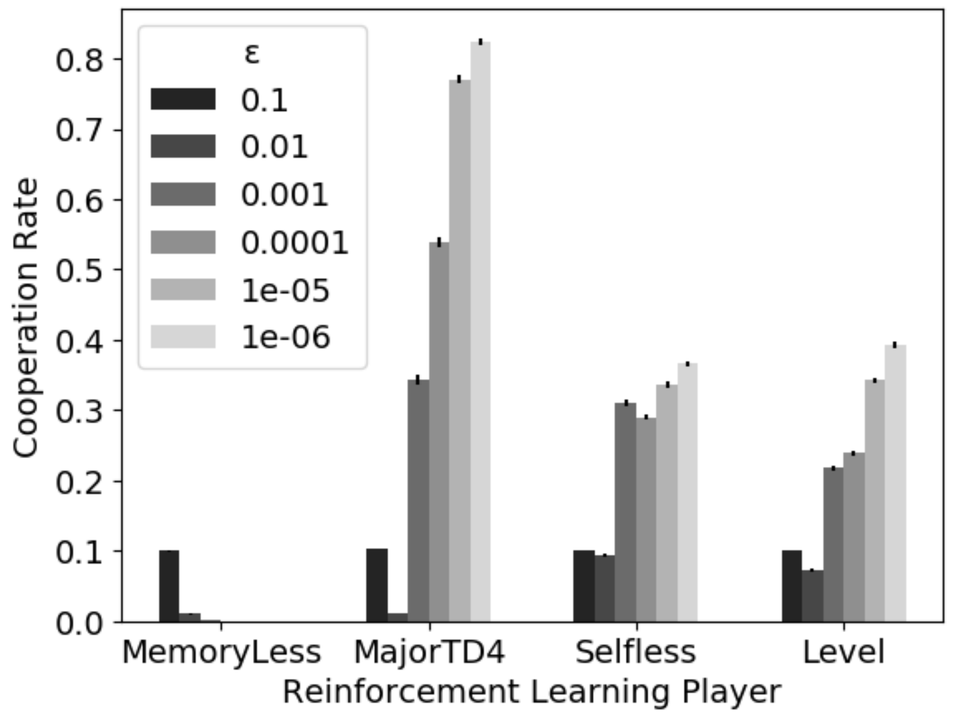

Figura 2: Taxas de cooperação de 5 agentes idênticos independentes com política de escolha de ações ε-greedy e diferentes espaços de estado.

Resultados e Conclusões

Para se aumentar cooperação deve-se aumentar tanto o nível de informação quanto melhorar o política de atuação, aumentar só um aspecto pode gerar ganhos limitados.

O agente que mais cooperou foi o *LevelLearner* com *Actor-Critic,* chegou a 80% de cooperação. Esta configuração aprendeu um mecanismo de recuperação que transita de um estado de baixa cooperação para um de alta cooperação rapidamente.

Estado	D0	C1	D1	C2	D2
P_{COOP}	0.58	0.18	0.06	0.24	0.02
Estado	**C3**	**D3**	**C4**	**D4**	**C5**
P_{COOP}	0.61	0.02	1	0.02	0.60

Tabela 1: Probabilidade média de se cooperar em cada estado aprendida por *LevelLearners* seguindo política *Actor-Critic* em *NPD*.

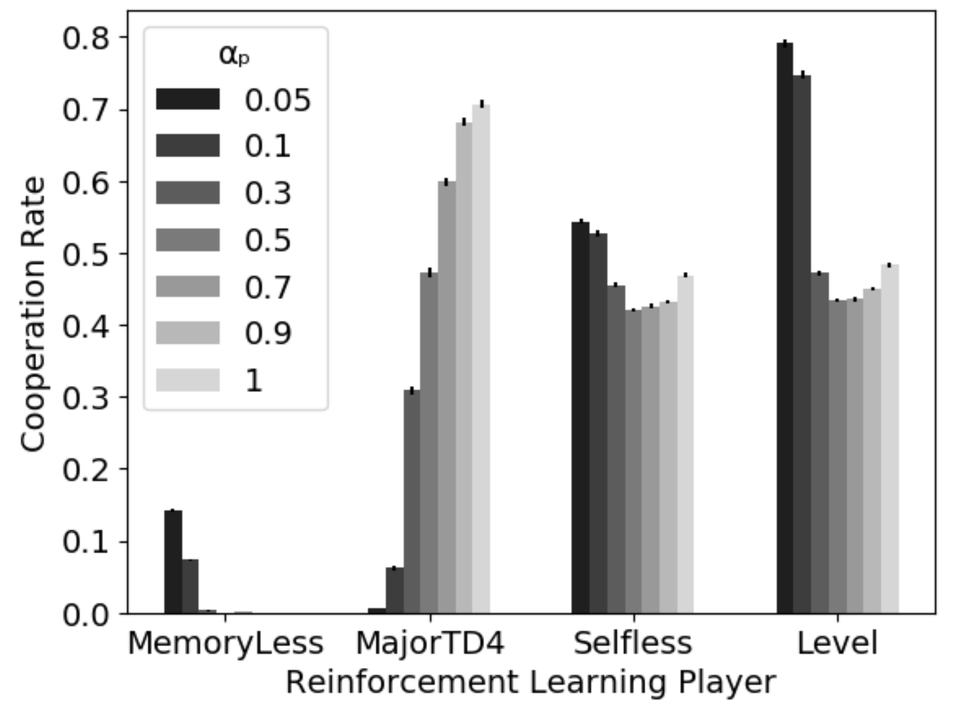

Figura 3: Taxas de cooperação de 5 agentes idênticos independentes com política de escolha de ações *Actor-Critic* e diferentes espaços de estado.

Referências

[1]: P. Kollock, “Social dilemmas: The anatomy of cooperation”, Annual review of sociology, vol. 24, no. 1, pp. 183–214, 1998.

[2]: N. Masuda and H. Ohtsuki, “A theoretical analysis of temporal difference learning in the iterated prisoner’s dilemma game”, Bulletin of Mathematical Biology, vol. 71, no. 8, pp. 1818–1850, Nov 2009.