

RELATO DE EXPERIÊNCIA

Controle de ângulo da junta do ombro do robô NAO com q-learning: Uma abordagem de aprendizado por reforço

Vitor Amadeu Souza¹*1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.*vitor.amadeu@foa.org.br[0009-0002-1857-6799](tel:0009-0002-1857-6799)

Resumo: O controle preciso das juntas de robôs humanoides, como as do ombro do robô NAO, é um desafio técnico devido à complexidade biomecânica e às não-linearidades dinâmicas, que frequentemente limitam a eficácia dos métodos tradicionais, como controladores PID. Este trabalho propõe uma abordagem baseada em aprendizado por reforço, utilizando o algoritmo Q-Learning, para controlar o ângulo da junta do ombro esquerdo do NAO, com o objetivo de alcançar um ângulo alvo de 45°. A metodologia foi desenvolvida e testada no ambiente de simulação CoppeliaSim, onde o problema foi modelado como um Processo de Decisão de Markov (MDP), com estados discretizados (ângulo e velocidade angular da junta) e ações representadas por torques aplicados. O algoritmo foi implementado em Python, utilizando a API ZMQ Remote do CoppeliaSim, e configurado com uma política ϵ -greedy para balancear exploração e aproveitamento. Os resultados mostraram que o Q-Learning convergiu em uma média de 62,70 episódios e desvio padrão de 2,72. Esses achados demonstram o potencial do Q-Learning como uma alternativa eficaz aos métodos tradicionais para o controle de juntas robóticas em simulação, contribuindo para o avanço de robôs humanoides em aplicações como assistência social e educação.

Palavras-chave: Robótica Humanoide. Política ϵ -greedy. CoppeliaSim. Python.

INTRODUÇÃO

A robótica humanoide tem avançado consideravelmente nas últimas décadas, com aplicações significativas em áreas como assistência social, educação e entretenimento (GOUAILLIER et al., 2009). No entanto, o controle preciso das juntas de um robô humanoide, especialmente as que envolvem o movimento dos membros superiores, continua sendo um desafio técnico complexo (KUINDERSMA et al., 2016). O robô NAO, desenvolvido pela Aldebaran Robotics (atualmente parte da SoftBank Robotics), destaca-se como uma plataforma amplamente utilizada para pesquisa em robótica humanoide devido à sua versatilidade e acessibilidade (SHAMSUDDIN *et al.*, 2011).

Em particular, as juntas do ombro do robô NAO apresentam desafios únicos devido à sua configuração biomecânica complexa, exigindo uma coordenação precisa com outras juntas para a execução de movimentos naturais (GONZÁLEZ-FIERRO et al., 2014). Métodos tradicionais de controle, como os controladores PID (Proporcional-Integral-Derivativo), frequentemente se mostram ineficazes ao lidarem com as não-linearidades e o acoplamento entre juntas típicos desse tipo de sistema (KOBBER; BAGNELL; PETERS, 2013).

Neste cenário, as técnicas de aprendizado por reforço (RL) têm se mostrado promissoras para o controle de sistemas robóticos complexos. O aprendizado por reforço oferece a vantagem de permitir que o agente aprenda a tomar decisões ótimas por meio da interação com o ambiente, sem a necessidade de um modelo explícito do sistema (SUTTON; BARTO, 2018). Dentre as abordagens de RL, o Q-Learning destaca-se por sua simplicidade conceitual e eficácia no enfrentamento de problemas de decisão sequenciais (WATKINS; DAYAN, 1992).

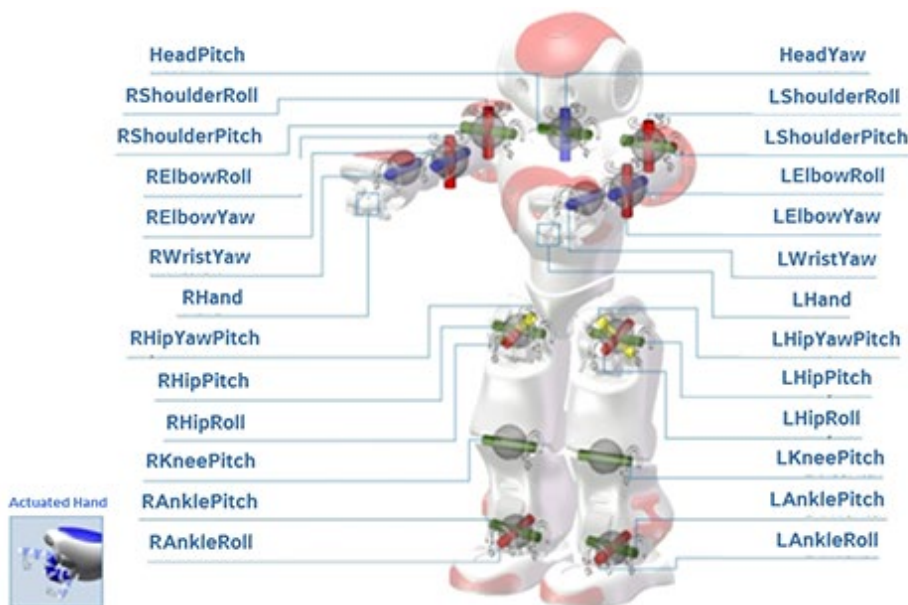
O presente trabalho propõe uma metodologia para o controle do ângulo da junta do ombro do robô NAO utilizando Q-Learning. A pesquisa se utiliza do ambiente de simulação CoppeliaSim (anteriormente conhecido como V-REP) para o desenvolvimento e testes iniciais do algoritmo. O objetivo principal é criar um controlador capaz de posicionar com alta precisão e eficiência a junta do ombro esquerdo do robô NAO em um ângulo de 45°, superando as limitações dos métodos convencionais como o PID.

METODOLOGIA

O NAO é um robô humanoide amplamente utilizado em pesquisa e educação (GOUAILLIER et al., 2009). Com 58 cm de altura e pesando aproximadamente 5,48 kg, o NAO possui 25 graus de liberdade (DOF), incluindo 2 DOF no pescoço, 5 DOF em cada braço, 1 DOF em cada mão, 5 DOF em cada perna e 1 DOF na pélvis (SHAMSUDDIN *et al.*, 2011), como mostra a Figura 1.

As juntas do ombro do NAO são particularmente importantes para a manipulação de objetos e para a expressão de linguagem corporal. Cada ombro possui 2 graus de liberdade (DOF): pitch (movimento para frente e para trás) e roll (movimento lateral). A rotação interna e externa do braço, conhecida como yaw, é realizada pela junta ElbowYaw, localizada no cotovelo. Estas juntas são acionadas por motores elétricos de corrente contínua equipados com encoders para feedback de posição (GOUAILLIER et al., 2009).

Figura 1 - RobÔ NAO com 25 DOF

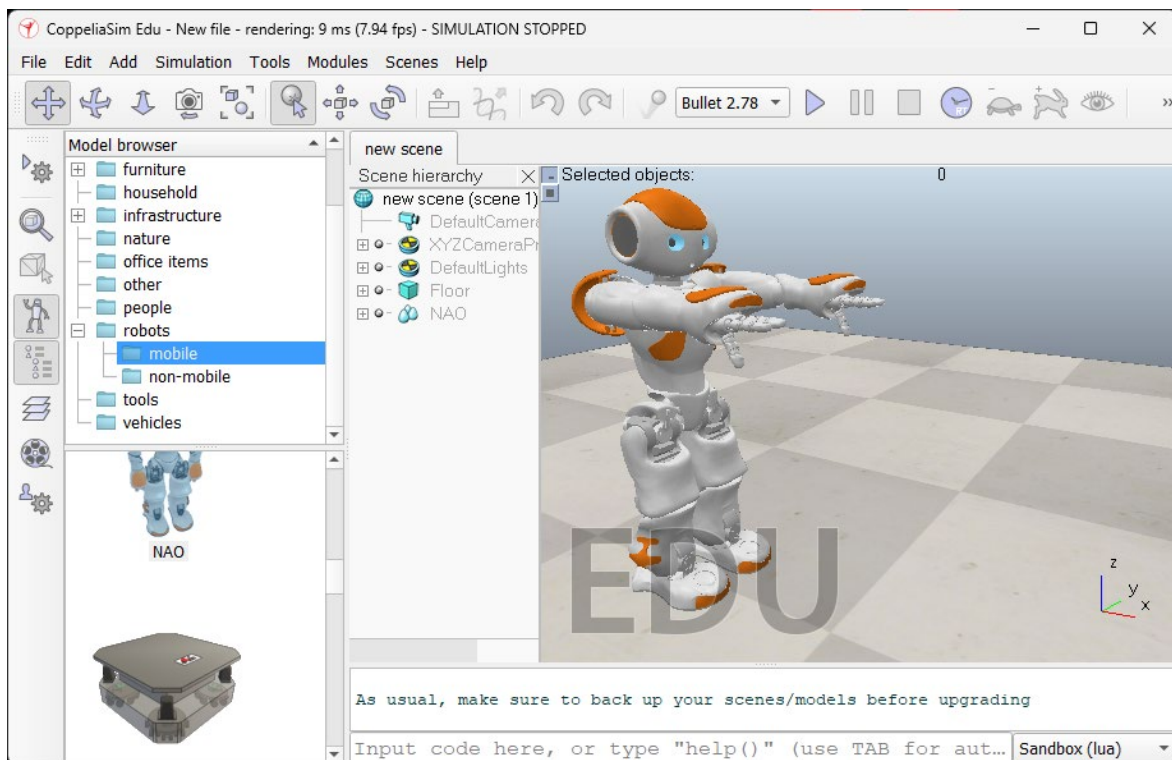


Fonte: (ALDEBARAN, 2025).

Este robô será simulado através do CoppeliaSim, que é um ambiente de simulação robótica que permite o desenvolvimento, teste e validação de algoritmos de controle antes da implementação em robôs físicos (ROHMER; SINGH; FREESE, 2013). O simulador oferece uma plataforma flexível com motor de física realista, permitindo simular as propriedades dinâmicas das juntas robóticas, como inércia, atrito e limites de torque. O CoppeliaSim suporta múltiplas APIs de programação, incluindo Python, o que facilita a implementação de algoritmos de aprendizado por reforço. Além disso, o simulador permite a criação de modelos precisos de robôs existentes, como o NAO, com representações fiéis das suas características físicas e cinemáticas. A Figura 2 mostra a interface deste simulador com o robô NAO alocado sobre o mesmo.

A utilização de simulação antes da implementação em robôs físicos é uma prática comum em robótica, pois permite o desenvolvimento e teste de algoritmos de forma segura e eficiente (KOBBER; BAGNELL; PETERS, 2013). No contexto do aprendizado por reforço, a simulação é particularmente valiosa, pois permite a coleta de grandes quantidades de dados de treinamento sem o desgaste do hardware físico.

Figura 2 - Ambiente do CoppeliaSim



Desta forma, o código-fonte em Python que roda no ambiente VS Code faz a comunicação com o CoppeliaSim através de uma API ZMQ Remote fornecida pelo fabricante, facilitando assim o processo de elaboração e depuração do algoritmo.

RESULTADOS E DISCUSSÃO

Para avaliar a eficiência do algoritmo Q-Learning aplicado ao controle do movimento do ombro esquerdo do robô NAO, foram realizados 10 ensaios. Cada ensaio consistiu na execução do algoritmo até que o ângulo-alvo de 45° fosse atingido ou até o limite máximo de 100 episódios. O número de episódios necessários para alcançar a convergência foi registrado em cada execução, conforme apresentado na Tabela 1.

Com base nesses dados, obteve-se uma média amostral de $x=62,70$ episódios. O desvio padrão amostral foi calculado como $\sigma = 2,72$, indicando uma dispersão relativamente baixa em torno da média.

Adicionalmente, foi determinado o intervalo de confiança de 95% para a média do número de episódios necessários até a convergência, utilizando a distribuição t de Student devido ao tamanho reduzido da amostra ($n = 10$). Dessa forma, pode-se afirmar, com 95% de confiança, que a média populacional do número de episódios necessários para que o robô alcance o ângulo desejado encontra-se entre 60,76 e 64,64.

A análise revela que o algoritmo apresentou desempenho estável, com variação limitada entre os ensaios. A baixa dispersão dos resultados e o intervalo de confiança estreito indicam que o agente é capaz de aprender uma política eficaz de maneira consistente. Portanto, sugere-se que o método proposto é confiável e eficiente para a tarefa de posicionamento do ombro do robô NAO utilizando aprendizado por reforço.

Tabela 1 - Quantidade de episódios necessários para alcançar o ângulo-alvo em cada ensaio

Ensaio	Episódio
1	61
2	59
3	65
4	61
5	64
6	61
7	66
8	64
9	67
10	59

Fonte: O Autor

O código-fonte completo, implementado em Python com integração via API está disponível para consulta no link (<https://github.com/vitor-souza-ime/qlearning>). Adicionalmente, diversos vídeos que demonstram o controle do braço do robô NAO utilizando o algoritmo Q-Learning podem ser acessados através do link (<https://tinyurl.com/5h9xu6xh>), oferecendo uma visualização prática dos resultados obtidos.

CONSIDERAÇÕES FINAIS

Este trabalho propôs uma abordagem de controle da junta do ombro do robô NAO utilizando Q-Learning, um algoritmo de aprendizado por reforço. A metodologia foi aplicada no ambiente de simulação CoppeliaSim, onde o controle do ângulo da junta do ombro esquerdo foi realizado com o objetivo de posicioná-la a 45°. A abordagem se mostrou promissora, superando as limitações dos controladores PID, especialmente quando se trata de lidar com não-linearidades e acoplamento entre as juntas do robô.

Os resultados experimentais indicaram que o Q-Learning foi capaz de aprender a política de controle de maneira eficaz, com um número médio de episódios para convergência de 62,70, o que demonstra que o algoritmo conseguiu encontrar uma solução eficiente para o problema de controle.

Com base nos resultados obtidos e nas limitações identificadas, diversas direções para trabalhos futuros podem ser exploradas. A validação do algoritmo em um robô NAO real constitui a principal continuação natural desta pesquisa, sendo essencial para avaliar sua eficácia em condições práticas, onde fatores como atrito, folgas mecânicas e incertezas nas medições podem influenciar significativamente o desempenho do controlador.

Adicionalmente, seria valioso realizar uma análise comparativa detalhada entre o controle baseado em Q-Learning e controladores tradicionais, estabelecendo métricas quantitativas para avaliar precisão, eficiência energética e robustez. A exploração de outros algoritmos de RL, como SARSA e Actor-Critic, também poderia revelar abordagens mais eficientes para o controle de juntas robóticas. A otimização sistemática dos hiperparâmetros do algoritmo Q-Learning e a investigação de técnicas de transfer learning para transferir conhecimento entre simulação e robô real representam caminhos promissores para aprimorar a eficiência do aprendizado.

REFERÊNCIAS

- GOUAILLIER, D. et al. Mechatronic design of NAO humanoid. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2009, Kobe. Anais... Piscataway: IEEE, 2009. p. 769–774.
- SHAMSUDDIN, S. et al. Humanoid robot NAO: Review of control and motion exploration. In: IEEE INTERNATIONAL CONFERENCE ON CONTROL SYSTEM, COMPUTING AND ENGINEERING, 2011, Penang. Anais... Piscataway: IEEE, 2011. p. 511–516.
- KUINDERSMA, S. et al. Optimization-based locomotion planning, estimation, and control design for Atlas. *Autonomous Robots*, v. 40, n. 3, p. 429–455, 2016.
- GONZÁLEZ-FIERRO, M. et al. Full-body postural control of a humanoid robot with both imitation learning and skill innovation. *International Journal of Humanoid Robotics*, v. 11, n. 2, p. 1450012, 2014.
- KOBER, J.; BAGNELL, J. A.; PETERS, J. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, v. 32, n. 11, p. 1238–1274, 2013.

SUTTON, R. S.; BARTO, A. G. Reinforcement learning: An introduction. Cambridge: MIT Press, 2018.

WATKINS, C. J.; DAYAN, P. Q-learning. Machine Learning, v. 8, n. 3-4, p. 279–292, 1992.

ALDEBARAN. Electronics architecture. Disponível em: http://doc.aldebaran.com/1-14/naoqi/sensors/dcm/low_level_architecture.html. Acesso em: 15 abr. 2025.

ROHMER, E.; SINGH, S. P.; FREESE, M. V-REP: A versatile and scalable robot simulation framework. In: IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS, 2013, Tóquio, Japão. Anais... [S.l.]: IEEE, 2013. p. 1321–1326.