Navegação via potência do sinal de rádio por meio de aprendizado por reforço

André Cid *

* *Departamento de Engenharia Elétrica, Universidade Federal de Minas Gerais, MG, (e-mail: andremacielcid@gmail.com).*

Abstract: Communication is one of the most critical areas in robot operations, especially in hazardous environments for humans, such as industrial environments. Maintaining a constant connection between the robot and the base ensures that mobile robot operations run as expected, reducing the risk of accidents, damage and the need to rescue the robot. Therefore, this work proposes a path planner that tries to maximize the connection between the robot and the transmitting antennas, ensuring that the robot moves to locations with the best radio signal. This is done using the Q-learning reinforcement learning algorithm and simulations of the radio signal strength in a discrete environment. Four methods for rewarding the robot based on radio signal strength are compared to analyze which of these methods performs best.

Resumo: A comunicação é uma das áreas mais críticas nas operações dos robôs, especialmente em ambientes perigos para humanos, tais como ambientes industriais. Manter uma conexão constante entre o robô e a base garante que operações com robôs móveis decorram como esperado, reduzindo o risco de acidentes, danos e a necessidade de resgate do robô. Assim, este trabalho propõe um planeador de caminhos que tenta maximizar a conexão entre robô e antenas transmissoras, garantindo que o robô se desloque por locais com o melhor sinal de rádio. Para isso, é utilizado o algoritmo de aprendizado por reforço Q-learning e simulações da potência do sinal de rádio em um ambiente discreto. São comparados 4 métodos para recompensar o robô com base na potência do sinal de rádio para analisar quais desses métodos apresentam melhor resultado.

Keywords: Reinforcement Learning; Comunication; navigation; Q-Learning.

Palavras-chaves: Aprendizado por reforço; Comunicação; Navegação; Q-Learning.

1. INTRODUÇÃO

E ambientes como os armazéns modernos, a eficiência operacional e a precisão são essenciais para manter a competitividade e atender às exigências crescentes do mercado. Uma das inovações mais impactantes nesse setor tem sido a implementação de robôs móveis, especialmente os robôs carregadores de cargas. Esses dispositivos automatizados são projetados para transportar e manejar materiais de forma autônoma, reduzindo o tempo de operação e minimizando os erros humanos. Contudo, a eficácia desses robôs depende fortemente da qualidade da comunicação constante entre eles e os sistemas de controle central.

A manutenção de uma conexão de comunicação robusta e constante não apenas potencializa a coordenação e a execução das tarefas, mas também assegura a segurança e a adaptabilidade em um ambiente que está sempre mudando.

Por exemplo, no caso dos robôs carregadores de cargas em armazéns, uma comunicação eficiente permite que eles naveguem eficazmente entre diferentes seções do armazém, otimizem as rotas de entrega e respondam dinamicamente a mudanças imprevistas no layout ou na demanda de armazenamento. Portanto, a conexão contínua de comunicação é fundamental não só para maximizar a eficiência operacional, mas também para garantir a integridade e a continuidade dos processos logísticos.

Dessa forma, uma alternativa para resolver o problema de comunicação é a navegação com base na comunicação, ou "communication aware navigation". Esse tipo de navegação representa um avanço significativo na robótica e nos sistemas autônomos, integrando a comunicação como um componente chave no processo de navegação. Esta abordagem foca em adaptar e otimizar a movimentação dos robôs em ambientes dinâmicos, levando em consideração a qualidade e a disponibilidade da comunicação sem fio. Em cenários onde múltiplos agentes autônomos, como drones ou robôs móveis, operam simultaneamente, a capacidade de ajustar suas trajetórias com base na conectividade de rede pode reduzir significativamente as interrupções e melhorar a eficiência coletiva.

Em uma aplicação prática, robôs equipados com tecnologias de navegação consciente da comunicação podem evitar áreas com sinal fraco, garantindo assim uma troca de dados constante e confiável, essencial para a execução coordenada de tarefas complexas e para a manutenção da segurança operacional. Este conceito não só melhora a autonomia dos robôs em termos de navegação espacial,

⋆ Reconhecimento do suporte financeiro deve vir nesta nota de rodapé.

mas também potencializa a robustez e a confiabilidade das operações robóticas em ambientes desafiadores.

Dessa forma, é proposto por este trabalho implementar um planejador de caminhos com base no algoritmo de Q-Learning para permitir que um robô móvel navegue em um ambiente considerado a comunicação como métrica importante. Para isso, são utilizadas 4 métricas para atribuir recompensas para o agente visando avaliar qual delas melhor representa um sistema de navegação com base na comunicação. Para o ambiente de teste, é utilizado um simulador para estimar a potência do sinal de comunicação. Por fim, são apresentados os gráficos de aprendizado e os mapas resultantes de cada método.

2. REVISÃO BIBLIOGRÁFICA

3. METODOLOGIA

Visando resolver o problema da navegação de um robô móvel em um local conhecido e sem ter o mapa da potência do sinal de rádio, é proposto a utilização de um método temporal (*Temporal-Difference Learning*) que utiliza da experiência de cada iteração para aprender. Dessa forma, foi escolhido o método Q-Learning. O método Q-learning é uma técnica de aprendizado por reforço que permite a um agente aprender a tomar decisões ótimas em um ambiente com base em recompensas. O objetivo do Q-learning é aprender uma política de ações que maximize a soma total das recompensas futuras. Outra vantagem no uso do Q-Learning é que ele é *model-free* e não requer um modelo do ambiente, ou seja, ele pode aprender diretamente da experiência. Esse ponto é importante para a navegação em um ambiente sem o mapa da potência do sinal de rádio conhecido pois a estimação dessa potência é complexa e depende de diversos fatores externos.

3.1 Agente e ações

A metodologia proposta para navegar em um ambiente de forma autônoma considera um robô terrestre com arquitetura diferencial, onde sua pose é representada por uma configuração (x, y, α) no espaço 2-D, onde α é a referência de rotação do robô. O robô é capaz de se locomover livremente pelo ambiente, e é equipado com um sensor capaz de medir a potência do sinal de rádio em um ponto do mapa. Seus estados podem ser descritos como a movimentação de um passo em todas as 8 direções ou ficar parado, totalizando nove possíveis estados.

3.2 Ambiente e estado

O método é desenvolvido para ser treinado em um ambiente virtual preparado com um modelo que estima/simula a potência do sinal de rádio em um ponto qualquer do mapa dado a posição das antenas do sistema. A grandeza da potência é dada em dB (Decibéis), e representada por dBm (Decibéis por milliwat) que corresponde à potência em decibéis do sinal em referência a 1 miliwatt. Essa métrica representa o valor da potência do sinal que chegará ao receptor de um sistema de comunicação dada as diferentes atenuações que o sinal pode sofrer durante o percurso do transmissor ao receptor.

Para a simulação, algumas premissas são utilizadas, sendo elas: o ambiente corresponde a um espaço de 10x10 metros, discretizado em um grid de 25x25; É pré-suposto que um boa qualidade da comunicação do sinal de rádio está diretamente ligada com a potência do seu sinal; Os valores de ganho das antenas e do transmissor são dados previamente assim como as posições das antenas transmissoras do sistema; As antenas transmissoras sem comportam de maneira igual.

Cada iteração da simulação tem como estados a posição do robô no grid discretizado e também a potência do sinal de rádio naquela posição. A equação que é utilizada para simular/estimar a potência do sinal de rádio é a *Log Distance Path Loss* (**?**). Essa equação (Eq. 1) associa a atenuação do sinal de rádio PL_L ao logaritmo da distância. Logo, dado uma atenuação conhecida PL_{d0}, em uma distância d_0 também conhecida, é possível calcular a atenuação do sinal de um ponto com distância d da antena transmissora considerando um fator n, que representa uma constante de propagação para cada ambiente em especifico, como o interior de prédios, com ou sem visada direta, fábricas, ambientes subterrâneo e etc.

$$PL_L = PL_{d0} + 10n \log \frac{d}{d_0}. \tag{1}$$

Logo, a potência P_r recebida por um transmissor pode ser encontrada dada a potência do sinal de saída P_t, os ganhos das antenas do transmissor e receptor(G_t, G_r e d, respectivamente) e a perda relacionada a propagação PL_L, conforme Eq. 2.

$$P_r(dbm) = P_t(dbm) + G_t + G_r + PL_{d0}. \tag{2}$$

Para estimar qual a potência do sinal de rádio na posição em que o robô se encontra, primeiro é calculado qual antena transmissora esta mais próxima. Para isso é utilizado a fórmula de distância entre dois pontos, conforme Eq. 3. A antena mais próxima é utilizada como base para calcular qual a potência do sinal dado a distância d do robô para essa antena.

$$d = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}. \tag{3}$$

Para essa simulação, foi implementado para cada antena transmissora com um sinal variando de -30dBm a -80dBm para os pontos mais e menos próximos respectivamente do grid do mapa. Também são considerada faixas para classificar a potência do sinal de rádio em "bom", "mediano" e "ruim". Os valores abaixo de -40dBm representam um bom sinal, uma faixa de sinal mediano fica entre -40dBm e -60dBm e valores menores que -60dBm representam sinal ruim.

Visualização A potência do sinal de rádio pode ser representada por meio de um mapa de calor, onde cada ponto é representado pela combinação de três cores (vermelho, azul e verde), para melhor visualização do comportamento do sinal no ambiente.

Dado um valor da potência do sinal P_{dBm}, primeiro é necessário normaliza-lo para um valor P_{rgb} entre 0 e 1. Essa normalização ocorre conforme Eq. 4, considerando o

valor ideal da potência do sinal de rádio dBm_{max} e o valor mínimo aceitável dBm_{min} para que ocorra a comunicação com o robô.

$$P_{rgb} = (P_{dBm} - dBm_{min}) \cdot \frac{1}{(dBm_{max} - dBm_{min})}, \quad (4)$$

O valor de cada faixa de cor é calculado utilizando as Equações 5 a 7 de forma que a cor vermelha representa maiores valores da potência do sinal de rádio e a cor azul menores valores.

$$Vermelho = P_{rgb}, \quad (5)$$

$$Verde = 0.8 \cdot (1 - 2 \cdot |P_{rgb} - 0.5|), \quad (6)$$

$$Azul = 1 - P_{rgb}. \quad (7)$$

3.3 Recompensas

Para testar a navegação do robô com base na potência do sinal de rádio, duas formas de avaliar a potência do sinal foram utilizadas como recompensa. Já para as recompensas relacionadas ao estado de posição do robô se mantiveram constantes

3.4 Recompensa - Posição

As recompensas relacionas à posição do robô foram:

- Se alcançar o objetivo, recebe +1000;
- Se o número de passo exceder 100, recebe -20;
- Se o robô colidir com algum obstáculo, recebe -50;

3.5 Recompensa - Sinal de rádio

Método 1

- Se sinal está maior que -40dBm, recebe +5;
- Se sinal está entre -40dBm e -60dBm, recebe +2;
- Se sinal está menor que -60dBm, recebe -5;

Método 2

- Se sinal está maior que -40dBm, recebe +5;
- Se sinal está menor que -60dBm, recebe -5;

Método 3

- Se sinal está menor que -60dBm, recebe -5;

Método 4

- Sinal é normalizado entre -5 e +5 conforme seu valor;

Por fim, é realizado o somatório de todas as recompensas para cada iteração.

3.6 Condições de fim

Cada simulação termina quando:

- O robô encosta em um obstáculo;
- O robô da mais de 100 passos;
- O robô chega no alvo;

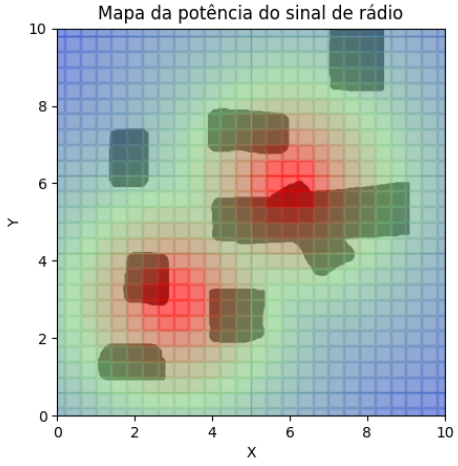

Figura 1. Mapa da potência do sinal de rádio com antenas posicionadas nas posições $x = 3, y = 3$ e $x = 6, y = 6$

3.7 Parâmetros

Para realizar o treinamento da rede, os seguintes parâmetros foram utilizados:

- Número de episódios: 200;
- Fator de desconto(Γ): 0.99;
- Epslon(ϵ): 0.02;
- Alpha (α): 0.5;

4. RESULTADO

O primeiro resultado esta relacionado ao mapa da potência do sinal de rádio do ambiente. Para isso, são utilizadas duas antenas, localizadas nas posições $x = 3, y = 3$ e $x = 6, y = 6$. O resultado da atenuação do sinal por todo o mapa pode ser visualizado na Fig. 1. Esse mapa é utilizado apenas para a visualização dado que os valores da potência do sinal de rádio não ficam disponíveis previamente para o robô, eles são calculados para cada iteração.

4.1 Métodos

Após utilizar as recompensas do primeiro método, o comportamento do sistema robótico pode ser visualizado na Fig. 2 e o gráfico de recompensa por episódio na Fig. 3. Os mesmo gráficos para o método 2, 3 e 4 podem ser vistos nas Figs. de 4 e 9.

Dados os mapas e gráficos acima, é possível notar que os métodos 2 e 3 foram os que apresentaram melhores resultados. Tal comportamento pode ser visto tanto nos movimentos do robô, que repelem áreas com baixa potencia do sinal de rádio e ainda conseguem chegar no objetivo.

5. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou alguns métodos possíveis para planejar um caminho em um ambiente discreto considerando a potência do sinal de rádio como métrica de avaliação. Para isso, primeiro foram apresentados as implementações realizadas em um simulador com o foco de representar a atenuação do sinal de rádio em um ambiente 2D. Em seguida, foram apresentas as configurações

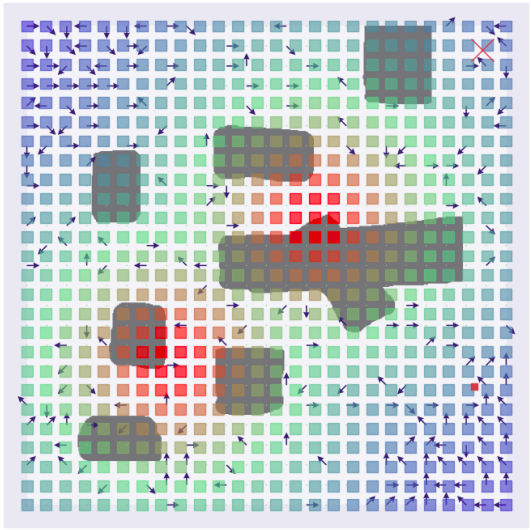

Figura 2. Comportamento do robô com método 1.

Figura 3. Gráfico de recompensa por episódio do método 1.

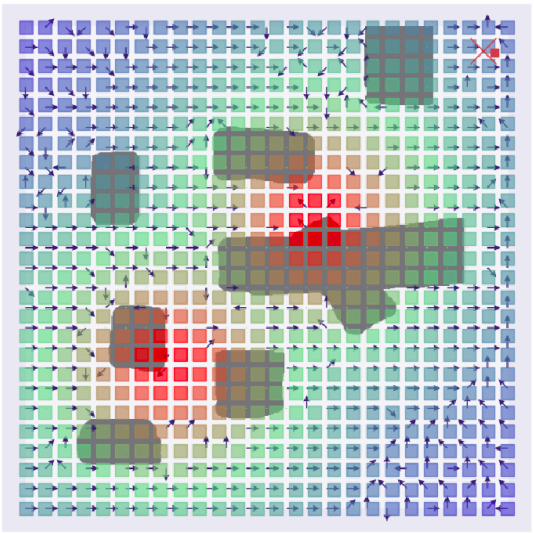

Figura 4. Comportamento do robô com método 2.

de parâmetros e recompensas utilizadas afim de conferir qual dessas melhor se encaixa no problema de navegar em um ambiente maximizando a conexão entre robô e antenas transmissoras. Todos os métodos foram testados no algoritmo Q-Learning.

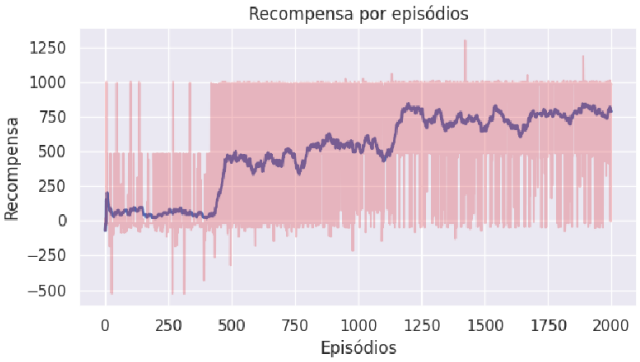

Figura 5. Gráfico de recompensa por episódio do método 2.

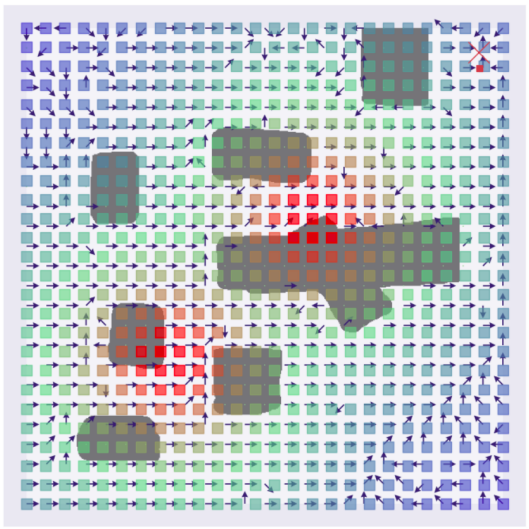

Figura 6. Comportamento do robô com método 3.

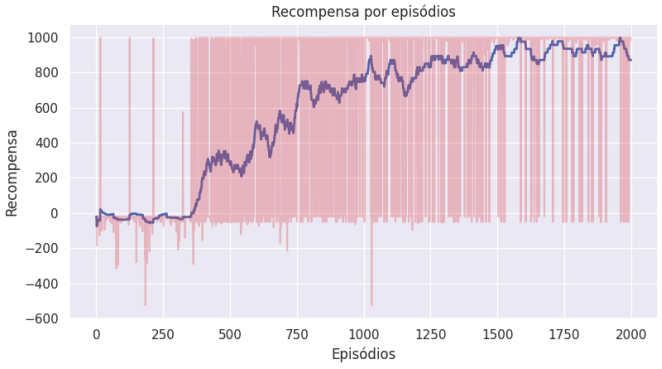

Figura 7. Gráfico de recompensa por episódio do método 3.

Os resultados demonstram que o método 3 foi o melhor para o dado objetivo uma vez que evitava passar em locais com baixa potência do sinal de rádio e também apresentou melhor curva de aprendizado.

Para trabalhos futuros, é proposto desenvolver um método de validação quantitavo afim de verificar qual método é realmente o melhor. Também é proposto o uso de métodos contínuos em um robô real para validar o uso de algoritmos

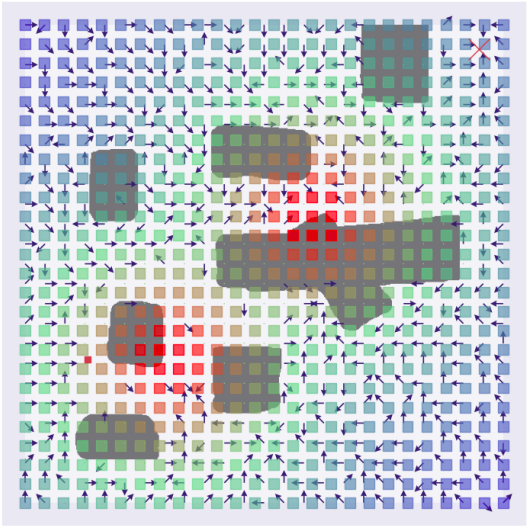

Figura 8. Comportamento do robô com método 4.

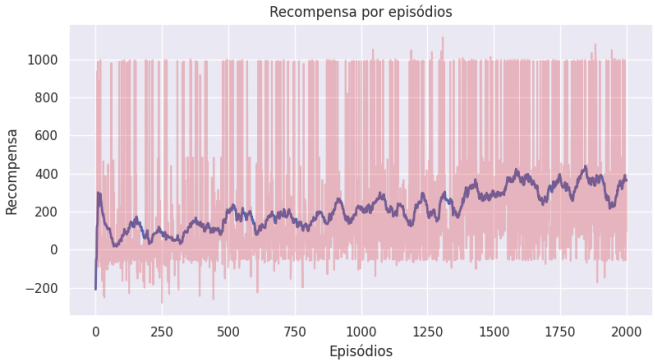

Figura 9. Gráfico de recompensa por episódio do método 4.

de aprendizado por reforço para navegação com base na potência do sinal de rádio.

REFERÊNCIAS