Universidade de Brasília

Faculdade de Tecnologia

Alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado

Breno de Almeida Beleza

PROJETO FINAL DE CURSO

ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Brasília

2025

Universidade de Brasília
Faculdade de Tecnologia

Alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado

Breno de Almeida Beleza

Projeto Final de Curso submetido como requisito parcial para obtenção do grau de Engenheiro de Controle e Automação.

Orientador: Prof. Dr. Paulo Henrique Portela de Carvalho
Coorientador: Prof. Dr. Paulo Roberto de Lira Gondim

Brasília
2025

FICHA CATALOGRÁFICA

de Almeida Beleza, Breno.

Alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado / Breno de Almeida Beleza; orientador Paulo Henrique Portela de Carvalho ; coorientador Paulo Roberto de Lira Gondim . -- Brasília, 2025.

50 p.

Projeto Final de Curso (Engenharia de Controle e Automação) -- Universidade de Brasília, 2025.

1. Aprendizado por Reforço Profundo. 2. JCAS. 3. Alocação de Potência. 4. D2D. 5. Aprendizado Federado. I. , Paulo Henrique Portela de Carvalho, orient. II. , Paulo Roberto de Lira Gondim, coorient. III. Título.

Universidade de Brasília
Faculdade de Tecnologia

Alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado

Breno de Almeida Beleza

Projeto Final de Curso submetido como requisito parcial para obtenção do grau de Engenheiro de Controle e Automação.

Trabalho aprovado. Brasília, 21 de fevereiro de 2025:

Prof. Dr. Paulo Henrique Portela de Carvalho,
UnB/FT/ENE
Orientador

Prof. Dr. Paulo Roberto de Lira Gondim,
UnB/FT/ENE
Coorientador

Prof. Dr. João Paulo Leite,
UnB/FT/ENE
Examinador interno

Me. Gabriel Pimenta de Freitas Cardoso
Examinador externo

Resumo

O avanço das redes sem fio e a crescente demanda por serviços de alta capacidade tornou essencial o desenvolvimento de tecnologias que permitam a reutilização do espectro para comunicação e sensoriamento do ambiente, dando origem aos sistemas de comunicação e sensoriamento integrados (JCAS, do inglês *joint communication and sensing*). Um dos principais desafios desses sistemas é a alocação eficiente de potência para garantir o equilíbrio entre a qualidade da comunicação e do sensoriamento. As técnicas de aprendizado de máquina (ML), em especial aprendizado por reforço profundo (DRL), surgem como alternativas promissoras para o controle de potência, devido a sua capacidade de adaptação a ambientes dinâmicos e complexos. Nesse contexto, outro desafio de grande preocupação é o de proteção e privacidade dos dados dos usuários, usados para o treinamento das redes DRL. Para lidar com esse problema, surgiu o paradigma de aprendizado federado (FL), que permite o treinamento colaborativo e local, no qual o modelo central é treinado a partir da agregação de parâmetros dos modelos locais, sem que haja necessidade de transmissão de dados dos usuários. Este trabalho propõe uma abordagem baseada em DRL federado para a alocação de potência em redes JCAS, analisado em três configurações diferentes de comunicação, explorando os algoritmos *Proximal Policy Optimization* (PPO) e REINFORCE, com o objetivo de avaliar o impacto da descentralização do aprendizado sobre a eficiência da alocação de potência, em termos da probabilidade de *outage* e da relação sinal ruído mais interferência (SNIR) das comunicações e dos sensores, comparando o desempenho dos modelos federados com as abordagens centralizadas tradicionais. Os resultados mostraram leve piora no desempenho da solução federada, quando comparada com a centralizada, porém sem expressividade suficiente para invalidar o emprego do FL, tendo em vista as vantagens de privacidade e segurança.

Palavras-chave: Aprendizado por Reforço Profundo; JCAS; Alocação de Potência; D2D; Aprendizado Federado.

Abstract

The advancement of wireless networks and the growing demand for high-capacity services have made it essential to develop technologies that enable spectrum reuse for both communication and environmental sensing, givind rise to joint communication and sensing (JCAS) systems. One of the main challenges of these systems is the efficient power allocation to ensure a balance between the quality of communication and the accuracy of sensing. Machine learning (ML) techniques, particularly deep reinforcement learning (DRL), have emerged as promising alternatives for power control due to their ability to adapt to dynamic and complex environments. In this context, another major concern is the protection and privacy of user data used to train DRL networks. To address this issue, the paradigm of federated learning (FL) has emerged, enabling collaborative and local training, in which the central model is trained by aggregating parameters from local models without requiring the transmission of user data. This work proposes a federated DRL-based approach for power allocation in JCAS networks, analyzed in three different communication configurations, exploring the Proximal Policy Optimization (PPO) and REINFORCE algorithms. The objective is to evaluate the impact of learning decentralization on power allocation efficiency in terms of communication and sensor outage probabilities and signal-to-noise-plus-interference ratio (SNIR), comparing the performance of federated models with traditional centralized approaches. The results showed a slight decrease in the performance of the federated solution compared to the centralized one; however, this degradation was not significant enough to invalidate the use of FL, considering its privacy and security advantages.

Keywords: Deep Reinforcement Learning; JCAS; Power Allocation; D2D; Federated Learning.

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

AWGN	Ruído Gaussiano Branco Aditivo
CAV	Veículo Automatizado e Conectado
CRLB	Limite inferior Cramér-Rao
CSI	Informação do Estado do Canal
DDPG	Deep Deterministic Policy Gradient
DDQN	*Dueling Deep Q Network*
DFRC	Função Dupla Radar-Comunicação
DL	Aprendizado Profundo
DNN	Rede Neural Profunda
DRL	Aprendizado por Reforço Profundo
FL	Aprendizado Federado
FTL	Aprendizado por Transferência Federado
HH	Hiper-Heurística
IA	Inteligência Artificial
IoT	Internet das Coisas
ISCC	Sensoriamento, Comunicação e Computação Integrados
JCAS	Comunicação e Sensoriamento Integrados
LOS	*Line-of-Sight*
MASAC	Soft Actor-Critic Multi-Agente
MDP	Processo de Decisão de Markov
ML	Aprendizado de Máquina
mmWave	Ondas Milimétricas
NO-T	Transmissão Não-Ortogonal
PL	*Path Loss*
PPO	*Proximal Policy Optimization*
PSO	Otimização por Enxame de Partículas
QoS	Qualidade de Serviço
RB	Bloco de Recursos
RCS	Seção Transversal do Radar
RL	Aprendizado por Reforço
SAC	Soft Actor-Critic
SF	*Shadow Fading Loss*
SGD	Descida Estocástica do Gradiente
SNIR	Relação Sinal Ruído Mais Interferência
UAV	Veículo Aéreo Não Tripulado

VoS Valor de Serviço

Sumário

1 Introdução

Com o avanço das redes sem fio e o crescimento exponencial da demanda por serviços de alta capacidade, tornou-se essencial desenvolver tecnologias que permitem uma melhor reutilização do espectro, suportando simultaneamente comunicações e sensoriamento do ambiente, reduzindo a necessidade de infraestrutura dedicada e melhorando a eficiência espectral (Fang *et al.*, 2022a). Assim surgiram os sistemas JCAS. Nesse contexto, a alocação eficiente de potência é fundamental para garantir um equilíbrio entre a qualidade da comunicação e a precisão do sensoriamento, especialmente em cenários que envolvem múltiplos usuários e dispositivos heterogêneos (González-Prelcic *et al.*, 2024). Desafios adicionais surgem quando, além das comunicações primárias e sensores, há diferentes tipos de comunicação, como dispositivo-para-dispositivo (D2D) (Cardoso, 2024).

Entre as abordagens para otimizar a alocação de potência em sistemas JCAS, o aprendizado por reforço profundo (DRL) tem se destacado por sua capacidade de adaptação a ambientes dinâmicos e complexos (Wang *et al.*, 2023). Além disso, abordagens descentralizadas, como o aprendizado federado (FL), vêm sendo exploradas para permitir o treinamento de modelos distribuídos, preservando a privacidade e segurança dos dados, reduzindo a necessidade de comunicação entre dispositivos (Lim *et al.*, 2020) e de conhecimento das informações do estado do canal, reduzindo a complexidade do sistema.

Este trabalho propõe uma abordagem baseada em aprendizado por reforço profundo federado para a alocação de potência em redes JCAS, explorando os algoritmos PPO (Proximal Policy Optimization) e REINFORCE. O objetivo é avaliar o impacto da descentralização do aprendizado sobre a eficiência da alocação de potência, comparando o desempenho dos modelos federados com as abordagens centralizadas tradicionais. O objetivo específico é avaliar o desempenho de técnicas DRL federadas em redes JCAS com três consumidores de banda, são eles: as comunicações primárias e D2Ds, e sensoriamento. Além disso, são considerados diferentes cenários de alocação de potência para analisar a robustez da solução proposta.

1.1 Ambiente

O estudo foi realizado em um espaço fabril de formato quadrado, constituído por um pátio central, onde se localizam a maior parte dos usuários, que são funcionários e máquinas, e um ambiente externo. Há, ainda, vias por onde trafegam os veículos, alvos do sensoriamento. Esse ambiente, mostrado na Figura 1.1, está situado no centro da célula de comunicação.

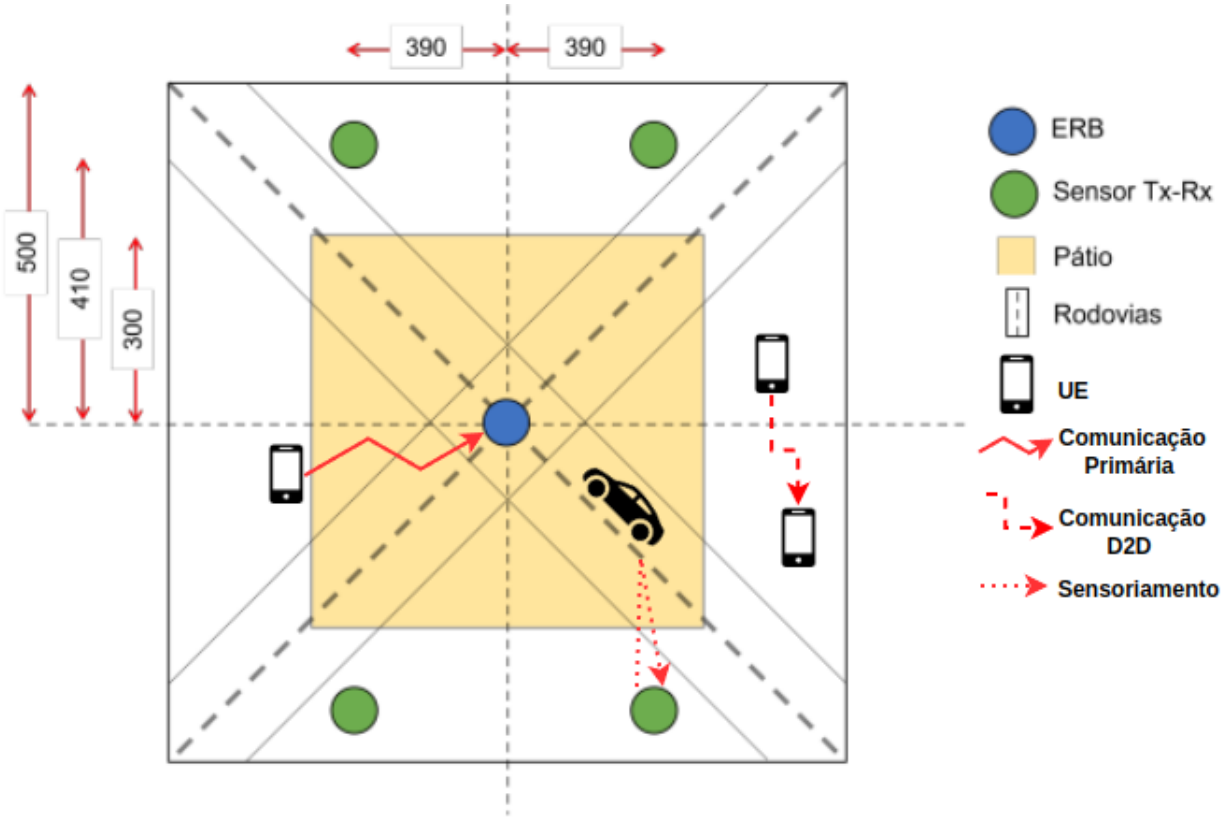

Figura 1.1 – Representação da célula modelada. Fonte: (Cardoso, 2024), com adaptações.

O ambiente será melhor detalhado no Capítulo 3, onde serão discutidos o canal de comunicação, os cenários de aplicação e a solução proposta.

1.2 Trabalhos Relacionados

Não foram achados, pelo autor, outros trabalhos que abordassem problemas de alocação de potência com os três consumidores de banda presentes em sistemas JCAS com D2D, portanto a revisão bibliográfica foi focada em artigos de alocação de potência para pares desses consumidores.

Em (Liu *et al.*, 2023), é analisado um sistema de comunicação e sensoriamento integrados (JCAS) onde a estação base fornece serviços de IoT e sensoriamento, utilizando transmissão não ortogonal (NO-T). O problema estudado é o de alocação de potência para maximizar as taxas de transmissão dos dispositivos de internet das coisas (IoT), respeitando os requisitos de sensoriamento. Um algoritmo alternativo de otimização é proposto, devido à não convexidade do problema. Segundo os autores, os resultados mostram que o sistema NO-T JCAS supera o JCAS convencional em relação às probabilidades de *Outage* e de detecção bem-sucedida, e o algoritmo proposto para alocação de potência supera os esquemas padrões, com maiores taxas de transmissão, enquanto respeita os requisitos de sensoriamento.

O estudo (Huang *et al.*, 2022) investiga uma rede JCAS que integra o canal de inter-

ferência para comunicação e sensoriamento distribuído por radar. Transmissores JCAS distribuídos enviam mensagens para usuários de comunicação enquanto cooperam para estimar a posição de um alvo. O controle de potência coordenado é explorado para minimizar a potência total dos transmissores, respeitando restrições de relação sinal-ruído mais interferência (SNIR) para comunicação e do limite inferior de Cramér-Rao (CRLB) (Taylor, 1978) para estimar a localização do alvo. Como o problema é não convexo, são propostos dois algoritmos baseados em relaxação semidefinida e aproximação de CRLB. Os resultados mostraram melhorias significativas na redução de potência em comparação com abordagens heurísticas.

Os autores de (Kim *et al.*, 2023) estudam um sistema *downlink* JCAS com um veículo aéreo não tripulado (UAV), com o objetivo de maximizar a taxa de transmissão enquanto garante a precisão do posicionamento de um alvo em movimento tridimensional, por meio da otimização da alocação de potência de transmissão. Foi proposto um modelo aproximado de transição de estado e um esquema de alocação de potência para o problema convexo resultante. Segundo os autores, os resultados numéricos mostram que o esquema proposto desempenha melhor que a *baseline* do treinamento de *beamforming* baseado em realimentação.

Já em (Qin *et al.*, 2023), os autores investigam o uso de múltiplos UAVs como plataformas JCAS móveis, otimizando comunicação e sensoriamento para usuários no solo. Foi formulado um problema integrado de associação de usuários, planejamento de trajetória e alocação de potência, maximizando a eficiência espectral mínima ponderada entre UAVs. Para lidar com a tomada de decisão sequencial, foram exploradas soluções de DRL centralizadas e descentralizadas, mais especificamente *soft actor-critic* (SAC) e SAC multi-agente (MASAC). Os resultados experimentais mostraram que o algoritmo SAC, centralizado, melhora a velocidade de treinamento e eficiência espectral, enquanto o MASAC, descentralizado, se destaca na velocidade inicial de treinamento.

O artigo (Chen *et al.*, 2024) estuda o problema de otimização da alocação de recursos *anti-jamming* em sistemas JCAS. O objetivo é maximizar a soma ponderada da taxa de comunicação e da potência de sensoriamento, enquanto garante os requisitos de comunicação e sensoriamento contra ataques do tipo *jamming*. Devido à complexidade do problema da integração entre comunicações e sensoriamento, além do comportamento dinâmico do *jamming*, foi proposto um algoritmo de aprendizado por reforço profundo (DRL) aliado à teoria dos jogos para a alocação de recursos. O controle de potência é modelado como um processo de decisão de Markov (MDP), e a seleção de canal como um jogo de Stackelberg (Nie; Zhang, 2008). Os resultados mostram ganhos significativos no desempenho da comunicação e sensoriamento no sistema JCAS e na resistência a interferências e ataques *jamming* .

O estudo (Cheng *et al.*, 2024) explora um modelo de patrulhamento dinâmico em redes JCAS assistidas por UAVs, que atuam como estações-base móveis para comunicação e unida-

des de radar e computação para sensoriamento. Para cumprir os requisitos de comunicação e sensoriamento, enquanto mitiga o impacto da mobilidade do UAV, foram propostos uma estrutura Mix-JCAS, que garante continuidade do sistema de comunicação, e um algoritmo DRL de otimização conjunta de trajetória e alocação de potência para melhorar a eficiência energética dos UAVs. Foi utilizada uma estratégia de amostragem multinível priorizada para aprimorar o treinamento, em comparação a estratégias tradicionais de amostragem não uniforme. Experimentos numéricos confirmaram a convergência e superioridade do algoritmo em comparação com abordagens convencionais.

Os autores em (Yang *et al.*, 2024a) investigam a alocação de potência em sistemas JCAS veiculares, considerando comunicação veículo-infraestrutura, veículo-veículo e sensoriamento, em meio a canais e tráfego dinâmicos. O problema é modelado como uma programação estocástica, que otimiza desempenho de sensoriamento, respeitando requisitos de estabilidade da rede e restrições de potência e qualidade de serviço. Utilizando otimização de Lyapunov, o problema é transformado em uma otimização não convexa, resolvida por um algoritmo híbrido que combina algoritmo genético e otimização por enxame de partículas (PSO). A análise teórica e os resultados da simulação mostram que a estratégia dinâmica proposta atinge um *tradeoff* entre o desempenho da comunicação e do sensoriamento.

O artigo (Liu *et al.*, 2024) propõe um esquema de alocação conjunta de canal e potência baseado em DRL para otimizar o uso de recursos em um sistema de internet de veículos com JCAS. Primeiramente, aplicam um algoritmo de política determinística profunda para alocar dinamicamente canais e potência com base na posição dos veículos. Depois, um algoritmo *deep Q-network* aloca banda de frequência para maximizar as taxas de comunicação. Os resultados de simulação mostram que o método proposto melhora a eficiência espectral e as taxas de comunicação em comparação com abordagens convencionais.

O estudo (Wang *et al.*, 2023) analisa um sistema JCAS baseado em OFDMA , onde estações base colaborativas transmitem sinais para usuários e realizam sensoriamento de múltiplos alvos. O objetivo é a alocação conjunta de subcanais e potência para maximizar a taxa total de transmissão, enquanto garante restrições de SINR para usuários e limite CRLB para alvos. Para resolver esse problema, é proposta uma abordagem baseada em DRL, utilizando *dueling deep Q network* (DDQN) e *deep deterministic policy gradient* (DDPG) para atribuição de sub canal e alocação de potência, separadamente. Os resultados demonstraram que a abordagem proposta aumentou a taxa de transmissão para todos os usuários, quando comparada com métodos padrões.

Os autores em (Zhao; Wu; Xiong, 2023) propõem um *framework* descentralizado de DRL multiagente para otimizar a percepção cooperativa em veículos automatizados e conectados (CAVs) usando JCAS. O problema de *beamforming* e alocação de potência é modelado como um MDP e depois expandido para um MDP parcialmente observável. A abordagem usa um sinal de *beacon* para transmissão de dados e uma forma de onda de função dupla radar-

comunicação (DFRC), que permite sensoriamento e comunicação V2V simultaneamente, sem necessidade de nós centrais. Resultados numéricos mostraram que o algoritmo obteve bom desempenho com rápida convergência, e forneceu qualidade de serviço (QoS) idêntica para todos CAVs.

O estudo (Yang *et al.*, 2024b) propõe uma arquitetura ISCC (sensoriamento, comunicação e computação integrados) para redes veiculares 6G, onde os veículos realizam sensoriamento do ambiente, computação de dados dos sensores, e transmissão. A abordagem utiliza aprendizado federado (FL) por computação *over-the-air* para permitir cooperação de baixa latência entre veículos e ampliar o alcance do sensoriamento. O problema de otimização conjunta de *beamforming* e alocação de potência é formulado para maximizar a taxa de dados enquanto mantém o desempenho de sensoriamento e computação. Para resolver esse desafio, o estudo propõe um esquema híbrido de DRL, combinando relaxação semidefinida, randomização gaussiana e DDPG. Simulações demonstraram melhor desempenho da abordagem em convergência e na maximização da taxa de dados, superando métodos convencionais.

O artigo (Noman *et al.*, 2024) propõe um método de DRL federado para gerenciamento eficiente de energia em redes heterogêneas assistidas por D2D. A abordagem utiliza uma *double-deep Q-network* para otimizar controle de potência e alocação dinâmica de canais, garantindo eficiência energética e QoS. Os resultados mostram uma melhoria na eficiência energética e um aumento na taxa de transmissão, quando comparada com outros algoritmos do estado-da-arte. Além disso, a taxa de *outage* celular foi reduzida, tornando o método robusto para redes 6G.

Os autores em (Alenezi; Luo; Min, 2021) propõem um algoritmo centralizado de controle de potência baseado em RL para otimizar a eficiência energética de dispositivos D2D em redes celulares, com o objetivo de reduzir o consumo de energia e minimizar a interferência nos usuários celulares, garantindo a QoS. Os resultados das simulações demonstram que a abordagem proposta aumenta a eficiência energética do sistema, mantendo a QoS dos usuários celulares em níveis superiores aos algoritmos de referência.

O estudo (Guo; Tang; Kato, 2022) propõe uma abordagem descentralizada baseada em DRL com FL para a alocação eficiente de recursos em redes sem fio habilitadas para D2D *underlay*, com o objetivo de maximizar a capacidade total e minimizar o consumo de energia, garantindo a QoS para usuários celulares e D2D. Segundo o autor, os resultados mostraram melhor desempenho, comparado com algoritmos de referência.

Por fim, o trabalho (Cardoso, 2024) propõe uma estratégia conjunta para alocação de espectro e controle de potência em redes 5G e futuras gerações com sensoriamento integrado, aplicada a cenários industriais da Indústria 4.0. A solução combina DRL e Hiper-Heurísticas (HH). O algoritmo de PPO mostrou melhor desempenho no controle de potência em um único bloco de recursos (RB), reduzindo a taxa de falha, probabilidade de *outage*,

das comunicações primárias de dos sensores, além de aumentar o SNIR das comunicações D2D. A estratégia completa (DRL+HH) obteve resultados semelhantes, mas para múltiplos RBs. A solução mostrou boa generalização para diferentes quantidades de comunicações, sensores e RBs.

As Tabelas 1.1 a 1.3 mostram as principais características dos estudos relacionados de maneira mais objetiva.

1.3 Contribuições deste trabalho

Este trabalho é uma continuação do estudo realizado em (Cardoso, 2024), focando na alocação de potência em redes JCAS com comunicações D2D utilizando algoritmos DRL, e possui as seguintes contribuições principais:

- Implementação do aprendizado federado aos algoritmos PPO e REINFORCE, comparando o desempenho ao centralizado;
- Análise de dois cenários adicionais, considerando apenas pares de consumidores de banda, são eles: comunicação primária e sensores, e comunicações primária e D2D.

1.4 Conclusão

Neste capítulo, foi discutida uma visão geral do contexto e da motivação para o estudo da alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado, destacando a importância da alocação eficiente de potência para otimizar a qualidade da comunicação e do sensoriamento.

Além disso, foram apresentados trabalhos relacionados que exploram diferentes abordagens para problemas similares, evidenciando a relevância do RL e do FL como estratégias promissoras para lidar com a complexidade e segurança dos dados dessas redes. Por fim, foram descritas as principais contribuições deste trabalho e a organização do restante do relatório.

Nos próximos capítulos, será abordada a fundamentação teórica necessária para a compreensão do estudo, depois a descrição do sistema, assim como dos problemas, cenários e a proposta de solução. Por fim, os resultados obtidos serão analisados, e o trabalho será concluído.

1.5 Organização do trabalho

O trabalho está estruturado da seguinte forma: o capítulo 2 aborda os conceitos fundamentais para o desenvolvimento do estudo, incluindo aprendizado DRL e FL. Além disso, resume os princípios de redes JCAS e a relevância da alocação de potência nesse contexto;

Artigo	Comunicação primária	Sensoriamento	Comunicação D2D	Aplicação	Problema	Objetivo	Solução
(Liu *et al.*, 2023)	✓	✓	-	IoT	Alocação da potência de transmissão	Maximização da taxa de transmissão	Algoritmo de otimização alternativa
(Huang *et al.*, 2022)	✓	✓	-	Comunicação e sensoriamento distribuído por radar	Alocação da potência de transmissão e estimação de posição	Minimização da potência total dos transmissores, com restrições de SNIR e CRLB	Algoritmos baseados em relaxação semidefinida e aproximação de CRLB
(Kim *et al.*, 2023)	✓	✓	-	UAV em movimento tridimensional	Alocação da potência de transmissão	Maximizar a taxa de transmissão e precisão do posicionamento do alvo	Aproximação de transição de estado para tornar o problema convexo, seguido pela alocação de potência
(Qin *et al.*, 2023)	✓	✓	-	UAVs como plataformas JCAS móveis	Alocação da potência de transmissão e planejamento de trajetória	Maximização da eficiência espectral	Algoritmos DRL centralizado e decentralizado, com arquiteturas SAC e MASAC
(Chen *et al.*, 2024)	✓	✓	-	Prenvenção de ataques a sistemas JCAS	Otimização de alocações *anti-jamming* de potência e canal	Maximização da taxa de comunicação e da potência de sensoriamento, respeitando requisitos contra ataques *jamming*	Controle de potência foi modelado como MDP e seleção de canal como jogo de Stackelberg
(Cheng *et al.*, 2024)	✓	✓	-	UAVs como EBs móveis	Otimização da alocação de potência e trajetória	Cumprir requisitos de comunicação e sensoriamento, enquanto mitiga impacto da mobilidade da UAV	Algoritmo DRL de otimização conjunta, com amostragem multinível priorizada
(Liu *et al.*, 2024)	✓	✓	-	IoV	Alocações de potência e canal	Maximização das taxas de comunicação	Algoritmos DDPG e DQN

Tabela 1.1 – Trabalhos relacionados, parte 1.

Artigo	Comunicação primária	Sensoriamento	Comunicação D2D	Aplicação	Problema	Objetivo	Solução
(Yang *et al.*, 2024a)	✓	✓		JCAS veiculares	Alocaçã o de potência	Otimizar desempenho de sensoriamento, com requisitos de estabilidade da rede e restrições de potência e QoS	Algoritmo genético e PSO
(Wang *et al.*, 2023)	✓	✓	-	JCAS multicelular	Alocação de potência e seleção de canal	Maximização da taxa total de transmissão, com restrições de SNIR e CRLB	Algoritmos DDQN e DDPG
(Zhao; Wu; Xiong, 2023)		✓		Percepção cooperativa de CAVs com JCAS	Alocação de potência e beamfor-ming	Maximização da QoS dos CAVs	DRL multiagente com conexões DFRC
(Yang *et al.*, 2024b)	✓	✓	-	Redes veiculares	Alocação de potência e beamfor-ming	Maximização da taxa de transmissão, mantendo desempenho do sensoriamento	FL com computação over-the-air, DDPG e relaxação semidefinida
(Noman *et al.*, 2024)	✓		✓	Redes hete-rogêneas	Alocação de potência e canal	Maximizar QoS e eficiência energética	Algoritmo DoubleDQN federado
(Alenezi; Luo; Min, 2021)	✓		✓	Redes celulares	Alocação de potência	Otimizar uso de energia e minimizar interferência	PPO

Tabela 1.2 – Trabalhos relacionados, parte 2.

Artigo	Comunicação primária	Sensoriamento	Comunicação D2D	Aplicação	Problema	Objetivo	Solução
(Guo; Tang; Kato, 2022)	✓		✓	Redes sem fio habilitadas para D2D *underlay*	Alocação de potência e canal	Maximizar capacidade e minimizar consumo de energia, enquanto garante QoS	DRL com FL
(Cardoso, 2024)	✓	✓	✓	Indústria 4.0	Alocações de potência e espectro	Redução de *Outage* da comunicação primária e sensores, enquanto maximiza a taxa de transmissão das comunicações D2D	Algoritmos DRL para alocação de potência e hiper-heurística para seleção de canal
Este Estudo	✓	✓	✓	Indústria 4.0	Alocação de potência	Redução de *Outage* da comunicação primária e sensores, enquanto maximiza a taxa de transmissão das comunicações D2D	Algoritmos DRL utilizando FL

Tabela 1.3 – Trabalhos relacionados, parte 3.

o capítulo 3 apresenta as modelagens do sistema de comunicação e dos problemas, e as descrições dos cenários analisados e da solução proposta; o capítulo 4 mostra, analisa e compara os resultados obtidos com a aplicação dos algoritmos de DRL, centralizados e federados, para a alocação de potências nos cenários propostos; e o capítulo 5 conclui o trabalho, que retoma o problema, a solução proposta e os principais resultados e análises.

2 Fundamentação teórica

2.1 Introdução

Este capítulo aborda conceitos essenciais para a compreensão da alocação de potência em redes JCAS com aprendizado por reforço profundo federado. Inicialmente, são apresentados os fundamentos de inteligência artificial e aprendizado de máquina, destacando suas principais abordagens, como aprendizados supervisionado e não supervisionado, e por reforço. Em seguida, são explorados conceitos de aprendizados profundo e por reforço profundo, e federado, essenciais para a implementação do modelo proposto. Por fim, são contextualizados sistemas de comunicação e sensoriamento integrados e a importância da alocação de potência para otimizar o desempenho dessas redes.

2.2 Inteligência artificial

Aprendizado de máquina (ML) é um ramo da inteligência artificial (IA) que permite que os computadores aprendam a realizar tarefas sem programação explícita, a partir de treinamento em conjuntos de dados que as representem (Eldar *et al.*, 2022).

O fluxo geral de aplicação de ML, ilustrado na Figura 2.1, começa com a obtenção dos dados, que, a depender da técnica empregada, podem ser divididos em conjuntos de treino e teste. Depois, os dados devem passar por tratamento, em preparação para o treinamento, mantendo as características mais importantes, e/ou criando novas (engenharia de características), para que as *features* e padrões possam ser extraídos e processados para a tarefa (Liu *et al.*, 2020). A seguir, há a seleção do algoritmo a ser empregado, seguido pelo treino, onde são aprendidos os padrões por meio de um mapeamento de entradas em saídas. O conjunto de teste, que são dados não antes vistos pelo algoritmo, passa por rotina semelhante, diferenciada no momento em que há a aplicação do modelo treinado, para enfim obter, visualizar e analisar os resultados.

Existem várias técnicas de ML, classificadas em 3 grupos principais, são eles: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço, que serão descritas nas seções seguintes.

2.2.1 Aprendizado supervisionado

O aprendizado supervisionado requer que o conjunto de dados para treino e teste seja rotulado, ou seja, contenha saídas correspondentes aos dados de entrada. A diferença entre a saída estimada e a real é utilizada ao longo do treinamento para medir a progressão

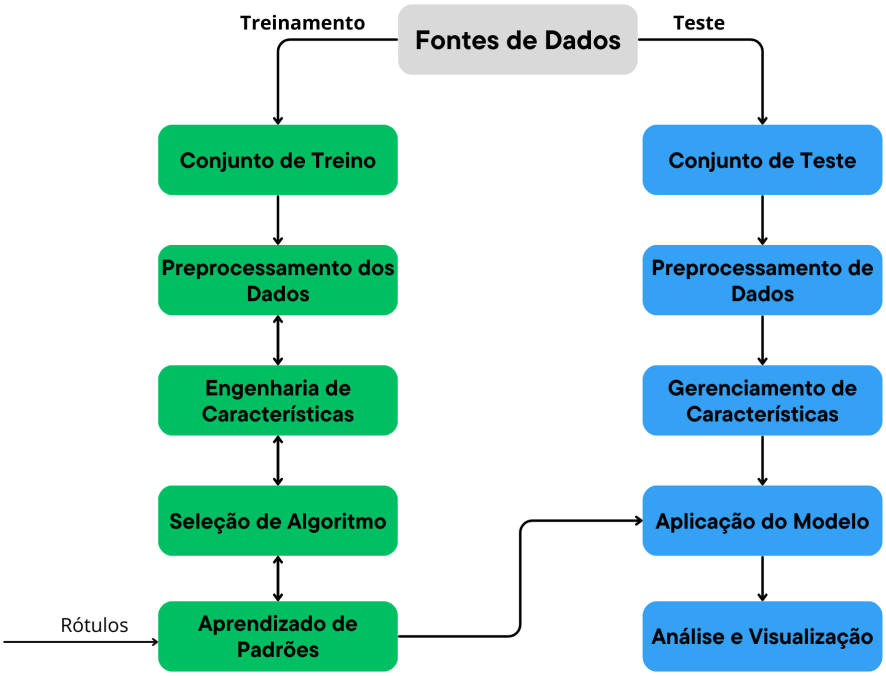

Figura 2.1 – Fluxo de uma aplicação em Machine Learning. Fonte: (Liu *et al.*, 2020), com adaptações.

do aprendizado a partir de uma função custo ou perda, e então são ajustados os pesos do algoritmo, visando reduzi-la (Engelbrecht, 2007). O objetivo do treinamento é a minimização da função, assim estimando uma função mapeamento, que na prática significa aprender a relação entre as entradas e saídas (Xie *et al.*, 2018), para fazer predições ou julgamentos que se aproximem da saída esperada. O modelo treinado é avaliado a partir dos dados de teste, desconhecidos para ele, a partir de métricas como acurácia, precisão, erro médio quadrado etc (Nithya *et al.*, 2023).

Aplicações incluem classificação de imagens, estimação de funções não lineares, entre outras. Alguns exemplos de algoritmos de aprendizado supervisionado são redes neurais, *Support Vector Machines* e árvores de decisão (Liu *et al.*, 2020).

2.2.2 Aprendizado não supervisionado

O aprendizado não supervisionado, em oposição ao supervisionado, recebe dados não rotulados, com o objetivo de descobrir características, padrões, estruturas e correlações (Nithya *et al.*, 2023) nos dados de entrada sem necessidade de caracterização prévia (Engelbrecht, 2007), com aplicações em agrupamento, detecção de anomalia, agregação de dados, entre outras (Liu *et al.*, 2020). Alguns exemplos de algoritmos de aprendizado não supervisionado são *Self-Organizing Feature Maps* e *K-means*.

2.2.3 Aprendizado por reforço

No aprendizado por reforço (RL), o treinamento é mais próximo do aprendizado humano e não requer grandes conjuntos de dados prévios (Uprety; Rawat, 2020), feito a partir da exploração do ambiente, usando mecanismos de recompensa e penalidade. Iterativamente, o agente, entidade do aprendizado (Xie *et al.*, 2018), interage com o ambiente, recebendo informações de estado e praticando ações, baseadas na estratégia que está sendo treinada e ajustada. As ações geram recompensas, pelas quais são avaliadas, e causam mudanças no estado do ambiente. Esse processo de aprendizado é repetido até que as condições de parada são alcançadas (objetivo atingido ou número máximo de iterações, por exemplo), assim finalizando o aprendizado (Zuo *et al.*, 2023). O objetivo geral do aprendizado por reforço é a maximização de uma recompensa de longo prazo, cumulativa e descontada, para que os agentes não só considerem a recompensa imediata, mas também o efeito das ações no futuro (Liu *et al.*, 2020).

O ambiente é descrito como um processo de decisão de Markov, que consiste em uma tupla $< S, A, T, R, \gamma >$, onde S é o conjunto de estados finitos, A é o conjunto de ações do agente, $T : S \times A \times S \rightarrow [0,1]$ é a função de probabilidade de transição, $R : S \times A \times S \rightarrow \mathbb{R}$ é a função recompensa, e $\gamma \in [0,1]$ é o fator de desconto, usado para balancear as recompensas imediatas e futuras (Li *et al.*, 2022). O estado $s_k \in S$ descreve o ambiente a cada passo de tempo discreto k. O agente pode modificar o estado ao executar ações $a_k \in A$. Como resultado da ação, o ambiente muda de s_k para algum s_{k+1}, de acordo com as probabilidades de transição de estado definidas em $T(s_k, a_k, s_{k+1})$, ou seja, a probabilidade de chegar ao estado s_{k+1} dado que a ação a_k foi executada no estado s_k. O agente recebe uma recompensa $r_{k+1} \in \mathbb{R}$, seguindo a função de recompensa $R : r_{k+1} = R(s_k, a_k, s_{k+1})$, que avalia o efeito imediato da ação a_k no estado s_k, que transiciona para s_{k+1}. Entretanto, não diz nada em relação ao efeitos a longo prazo. O comportamento do agente é descrito pela sua política π, que especifica como a ação é escolhida, dado o estado. Ela pode ser estocástica, $\pi : S \times A \rightarrow [0,1]$, definindo probabilidades para a escolha de ações em um estado, ou determinística, $\pi : S \rightarrow A$, com probabilidade 1 de ser escolhida uma ação a_k dado s_k (Busoniu; Babuska; Schutter, 2008).

O objetivo do agente é, portanto, achar uma política ótima π^* que maximize a recompensa descontada esperada, que é a recompensa esperada acumulada pelo agente no longo prazo, definido como:

$$R_k = E\{\Sigma_{k=0}^{\infty} \gamma r_k(s_k, a_k)\}, a_k \sim \pi(s_k), \tag{2.1}$$

onde E é o operador de valor esperado, ou esperança, e $\sim$ representa que o termo à esquerda se comporta como política $\pi(s_k)$.

Uma maneira de conseguir otimizar a política é com a função ação-valor (Função Q), que representa o retorno esperado de um par estado-ação dada a política π (Busoniu; Babuska; Schutter, 2008):

$$Q^{\pi}(s,a) = E\{\Sigma_{j=0}^{\infty}\gamma^{j} r_{k+j+i} | s_k = s, a_k = a, \pi\} \tag{2.2}$$

onde k é um passo de tempo discreto, e i

A função ótima é definida como $Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a)$, e satisfaz a equação de Bellman:

$$Q^*(s,a) = \Sigma_{s' \in S} T(s,a,s')[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')] \tag{2.3}$$

A Equação 2.3 diz que o valor ótimo da ação a em s é a soma das recompensas imediatas $R(s,a,s')$ esperadas somadas aos valores ótimos esperados descontados para todos os possíveis próximos estados s', multiplicados pelas respectivas probabilidades $T(s,a,s')$ de ocorrerem.

2.2.4 Aprendizado profundo

Algoritmos convencionais de ML dependem de extratores de características manualmente implementados, sendo necessário customização e reinicialização para cada novo problema. Com o aprendizado profundo (DL), por outro lado, as redes neurais automaticamente aprendem as características a partir dos dados brutos. Redes de DL frequentemente apresentam desempenhos melhores que algoritmos de ML (Surakhi *et al.*, 2021), principalmente quando há abundância de dados (Lim *et al.*, 2020).

Essencialmente uma rede neural é composta de 3 tipos de camadas, são elas: a camada de entrada, a oculta, e a de saída. Uma rede genérica é mostrada na Figura 2.2:

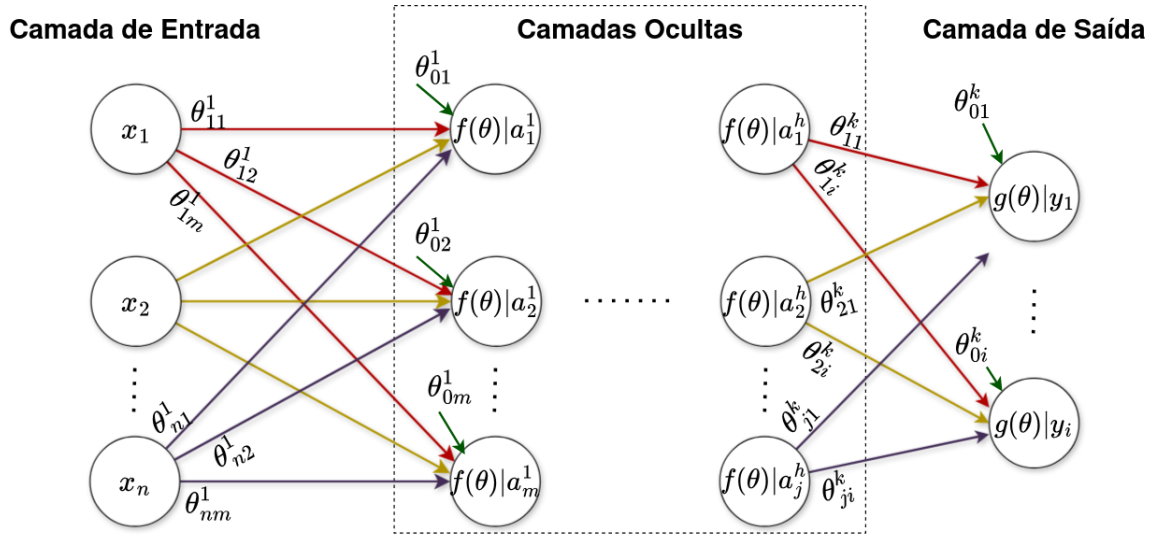

Figura 2.2 – Diagrama de uma rede neural. Fonte: Autor

Na rede representada na Figura 2.2 há k camadas: a inicial com n neurônios de entrada $X = \{x_1, x_2, ..., x_n\}$; h camadas ocultas de quantidade variável de neurônios, que geram as saídas intermediárias $A = \{\{a^1_1, ..., a^1_m\}, ..., \{a^h_1, ..., a^h_j\}\}$; e a camada de saída, que gera i

saídas $Y = \{y_1, ..., y_i\}$. Os valores de entrada são multiplicados por pesos $\theta^k_{\#*}$, corrigidos por *biases* θ^k_{0*}, onde # e $*$ são os neurônios de origem e de destino, respectivamente, e depois passados por uma função de ativação $f(\theta)$ para gerar saídas intermediárias a^h, que serão as entradas da próxima camada, até a última, que gera a saída da rede, a partir de uma função de ativação $g(\theta)$ que pode ser diferente das outras.

2.2.4.1 Fluxo de DL

Conforme mencionado, antes do treinamento, os dados são divididos em conjuntos de treinamento e de teste. O primeiro é utilizado como um conjunto de entradas, e saídas no caso de aprendizado supervisionado, para otimização dos pesos da rede, a partir de um algoritmo de otimização. Um exemplo de algoritmo é o descida estocástica do gradiente (SGD), no qual o conjunto de pesos θ é atualizado pelo produto da taxa de aprendizado ε, que representa o tamanho do passo do algoritmo a cada iteração, com a média aritmética dos gradientes da função perda L em relação a θ em um *minibatch* de tamanho m' (Goodfellow; Bengio; Courville, 2016), como segue:

$$\theta \leftarrow \theta - \varepsilon g, \tag{2.4}$$

$$g = \frac{1}{m'} \nabla_\theta \Sigma_{i=1}^{m'} L(x^i, y^i, \theta) \tag{2.5}$$

As iterações de treinamento são repetidas ao longo de várias épocas, cada uma sendo uma passagem completa pelo conjunto de treinamento. Uma rede de DL bem treinada deve ter capacidade de generalização, ou seja, conseguir realizar inferência de dados com os quais a rede não foi treinada, como o conjunto de testes, com certa acurácia (Lim *et al.*, 2020).

2.2.5 Aprendizado por reforço profundo

Técnicas tradicionais de RL guardam as informações do treinamento, como estado, ação e recompensas, de forma tabular, que não é escalável para um grande número de interações, ou para ambientes com estado-ação contínuo. Umas das soluções para isso é o DRL, que consiste em aplicar redes neurais profundas (DNNs) aliadas ao RL (Munikoti *et al.*, 2023), utilizando sua capacidade de representação para processar dados de grandes dimensões e aproximar os valores de estado-ação. Assim, a DNN mapeia os estados originais em *features*, que são mapeadas em ações (Wang *et al.*, 2022) .

Além da redução de dimensionalidade para os valores estado-ação, o DRL tem a vantagem de generalizar o valor de estados para os quais não foi treinado, a partir de estados similares, o que permite aplicação em ambientes complexos, com grande espaço de estados. Ademais, a DNN pode ser usada para estimar a recompensa de seguir uma certa política (Munikoti *et al.*, 2023).

Um dos métodos possíveis para DRL é o baseado em política, que aprende diretamente a política, em vez de focar nos valores para depois achar a política otimizada. Dois algoritmos de referência baseados em política são REINFORCE (Williams, 1992) e *Proximal Policy Optimization* (PPO) (Schulman *et al.*, 2017).

2.2.6 Aprendizado federado

No ML tradicional o treino é feito de forma centralizada, ou seja, com dados concentrados em um único servidor. Entretanto, a crescente preocupação com a privacidade dos dados de treinamento que pertencem a usuários de um sistema (Li *et al.*, 2021) deu origem ao conceito de aprendizado federado (FL) (Lim *et al.*, 2020).

O FL é um paradigma de ML que permite que os modelos sejam treinados de forma colaborativa, paralela e local nas fontes de dados, sem que haja a transmissão de dados locais. No FL, o modelo global é sintetizado à partir das atualizações de modelos locais, treinados nos dispositivos de borda, com seus próprios dados. Uma vantagem imediata é permitir que o modelo seja treinado com dados que não poderiam ser usados no modelo centralizado, devido a problemas de privacidade e eficiência (Jeno, 2022), além de aumentar a segurança, já que os dados dos usuários não saem de suas máquinas (Bhatia; Samet, 2022). Por outro lado, o FL apresenta um desempenho menor para aprendizado de características quando comparado com o treinamento clássico, uma vez que o modelo global não aprende diretamente das características dos dados (Alsaleh; Menai; Al-Ahmadi, 2024).

2.2.6.1 Fluxo do FL

Existem duas principais entidades no sistema FL, são elas: os donos dos dados, usuários, e o do modelo, servidor.

O processo de FL é iterativo, constituído por etapas que se repetem a cada iteração, ilustradas na Figura 2.3 e descritas a seguir (Jeno, 2022):

1. O modelo global mais atual é enviado para cada dispositivo de usuário.
2. Os modelos localizados em cada usuário são treinados com dados locais.
3. Depois de uma quantidade de rodadas de treinamento os parâmetros dos modelos locais são enviados ao servidor central.
4. O servidor central agrega os modelos locais a partir de uma função de agregação, gerando um novo modelo global agregado, que vira referência para a próxima iteração. Uma maneira de implementar é a partir de uma média ponderada pelo tamanho do conjunto de dados locais (Das *et al.*, 2024), como mostrado na Equação 2.6.

$$\theta_g = \frac{1}{D} \sum_{k \in K} D_k \, \theta_k, \tag{2.6}$$

onde, θ_g e θ_k são os parâmetros dos modelos global e local do dispositivo $k \in K$, respectivamente, e D e D_k são as quantidade de dados total do sistema e local do dispositivo.

Esta maneira não funciona para todos os tipos de FL, conforme descrito na próxima seção.

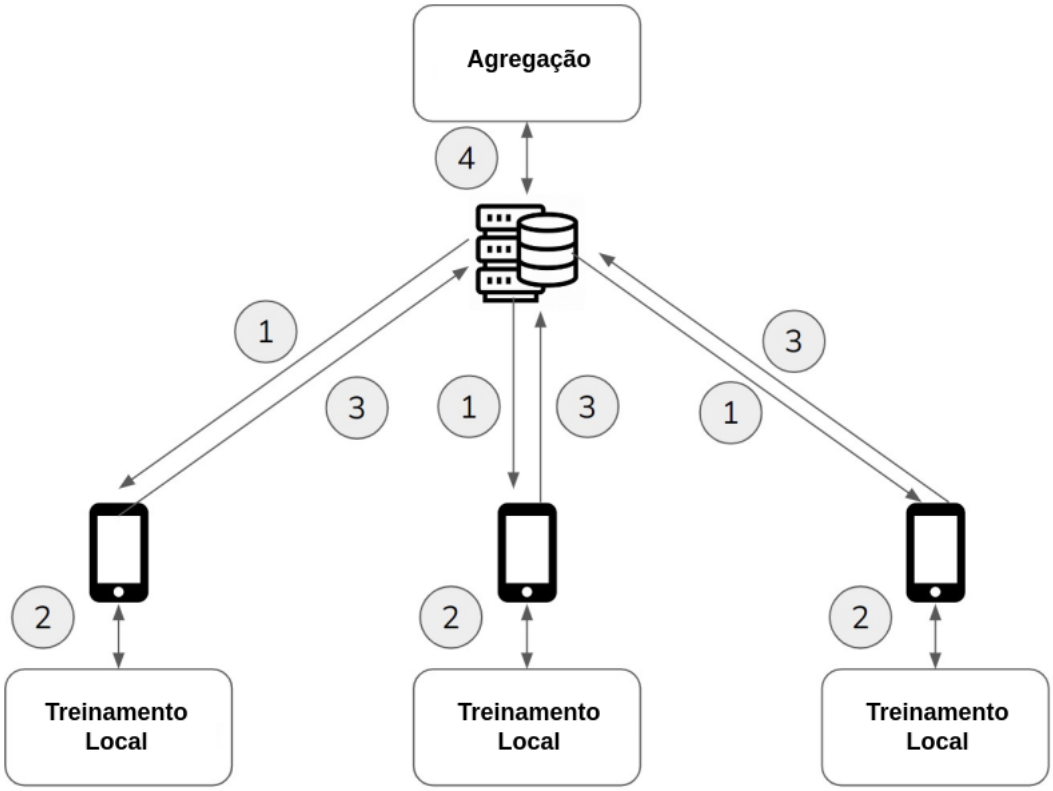

Figura 2.3 – Etapas de uma iteração de um sistema de FL. Fonte: (Jeno, 2022)

2.2.6.2 Tipos de FL

Os sistemas FL podem ser classificados de acordo com as características de distribuição dos dados, e tipicamente são de 3 tipos: horizontal (ou baseado em amostras), vertical (baseado em *feature*) e aprendizado por transferência federado (FTL) (Al-Quraan *et al.*, 2021)

1. FL horizontal: Categoria mais comum, na qual os conjuntos de dados de diferentes entidades têm mesmo espaço de *features*, mas espaços amostrais diferentes. Em outras palavras, têm mesmo objetivo de aprendizado, mas com tipos de dados diferentes. Esta categoria facilita o emprego de um modelo ML unificado, com mesma arquitetura para todos os conjuntos de dados.
2. FL vertical: Esta categoria pode ser empregada quando os conjuntos têm mesmo espaço amostral, mas espaços de *features* diferentes. Em outras palavras, possuem dados de mesma origem, mas têm objetivos de aprendizado diferentes.
3. FTL: Também chamado de aprendizado híbrido, permite transferência de conhecimento de um domínio para o outro, e pode ser usado quando há interseções entre

os espaços amostrais e de *features* de clientes diferentes. Um exemplo comum é o de classificação de imagens, onde modelos treinados para conjuntos de dados específicos podem ser usados para classificação de outros tipos de dados, após pequenos ajustes.

2.3 Sistemas de comunicação e sensoriamento integrados

JCAS é um paradigma de redes sem fio que integra comunicação e sensoriamento em um sistema unificado, permitindo maior eficiência espectral, o que o torna uma tecnologia chave para redes 6G (Fang *et al.*, 2022a).

Sistemas JCAS compartilham recursos de processamento e de *hardware* para suportar a integração, o que pode otimizar o desempenho, reduzir custos, e minimizar requisitos de infraestrutura (González-Prelcic *et al.*, 2024), além de benefícios mútuos para a comunicação e o sensoriamento, como melhor coordenação entre os múltiplos nós do sensoriamento, e maior reconhecimento do ambiente para as comunicações, com potencial para aumento de segurança e desempenho (Zhang *et al.*, 2021).

2.3.1 Tipos de sistemas JCAS

Os sistemas JCAS podem ser classificados em 3 tipos de configurações, são elas:

1. Radar-cêntrico: Esses sistemas integram comunicações em um sistema de sensoriamento priorizando a funcionalidade do radar, focando em realçar as capacidades de sensoriamento, sem alterar significativamente sua infraestrutura.
2. Comunicação-cêntrico: Nessa abordagem, a capacidade de sensoriamento é incorporada em sistemas de comunicação primária, enfatizando a integração do sensoriamento em redes móveis visando comunicação e sensoriamento contínuos.
3. Otimização e *design* integrados: Esse tipo de sistema procura otimizar a comunicação e o sensoriamento conjuntamente, provendo maior flexibilidade em termos de *design* da forma de onda, processamento de sinal, e arquitetura do sistema para alcançar performance superior para ambas as funcionalidades.

2.3.2 Alocação de potência

Um dos fatores críticos em sistemas JCAS é a alocação dinâmica de potência entre as comunicações primárias e os sensores, que gerencia soluções de compromisso entre as comunicações e o sensoriamento. O controle eficiente da potência é importante para garantir a qualidade dos mesmos.

As estratégias típicas de alocação de potência de sistemas JCAS incluem abordagens baseadas em otimização que consideram fatores como QoS e SNIR, e algoritmos de ML para

predizer distribuições ótimas de potências baseadas em dados históricos e estados atuais da rede (González-Prelcic *et al.*, 2024).

2.4 Conclusão

Neste capítulo, foram brevemente apresentados os conceitos fundamentais para o desenvolvimento deste trabalho, incluindo técnicas de aprendizado de máquina, focando nos paradigmas de aprendizado por reforço, profundo e federado, o conceito de sistemas JCAS e alocação de potência nesse contexto.

3 Aplicação de aprendizado por reforço profundo federado para alocação de potência em sistemas JCAS

3.1 Introdução

A alocação de potência é fundamental para eficiência de serviços de redes 5G e 6G, uma vez que a potência de transmissão afeta a qualidade do sinal e da comunicação e, portanto, deve ser ajustada para manter a qualidade do sinal de acordo com as interferências e questões ambientais, aumentar *throughput*, garantir acurácia do sensoriamento etc (Su; Lübke; Franchi, 2024).

Neste capítulo, é apresentada a aplicação do aprendizado por reforço profundo federado para a alocação de potência em sistemas JCAS. Inicialmente, descreve-se a estrutura do sistema de comunicações modelado em (Cardoso, 2024), detalhando os tipos de conexões envolvidas, os elementos presentes na rede e os parâmetros que influenciam a transmissão de dados e o sensoriamento. Em seguida, são discutidos os desafios relacionados à alocação dinâmica de potência, considerando a coexistência de comunicações primárias, comunicações D2D e sensores em um ambiente compartilhado. Três cenários distintos são propostos para avaliar a alocação de potência sob diferentes condições. Por fim, é apresentado o modelo de aprendizado por reforço profundo federado, descrevendo sua formulação matemática e implementação para solucionar o problema de alocação de potência, garantindo eficiência e equilíbrio entre comunicação e sensoriamento.

3.2 Sistema de comunicações

O sistema de comunicações modelado consiste em uma única célula e bloco de recurso, com diferentes tipos de comunicações. São realizadas até $J = \{1, 2, ..., 6\} = A + L$ comunicações, formadas por $A = 1$ comunicação primária no *uplink*, que ocorrem entre um UE e uma ERB, e $L = \{1, 2, ..., 5\}$ comunicações D2D, diretas entre os dispositivos na mesma faixa de frequência. O sistema possui $Q = \{1, 2, 3, 4\}$ sensores para sensoreamento de veículos.

A célula modelada, que possui um formato quadrado de lados de 1 km, com a ERB no centro, é mostrada na Figura 3.1.

onde g_A, g_l e g_q^e são os ganhos de propagação dos canais direto das comunicações primária A e D2D l, e com reflexão no alvo do sensor q, respectivamente, e representam a situação do canal. Além disso, $h_{l,A}$, $h_{l,q}$ e $h_{q,A}$ são: os ganhos de propagação dos canais entre o transmissor

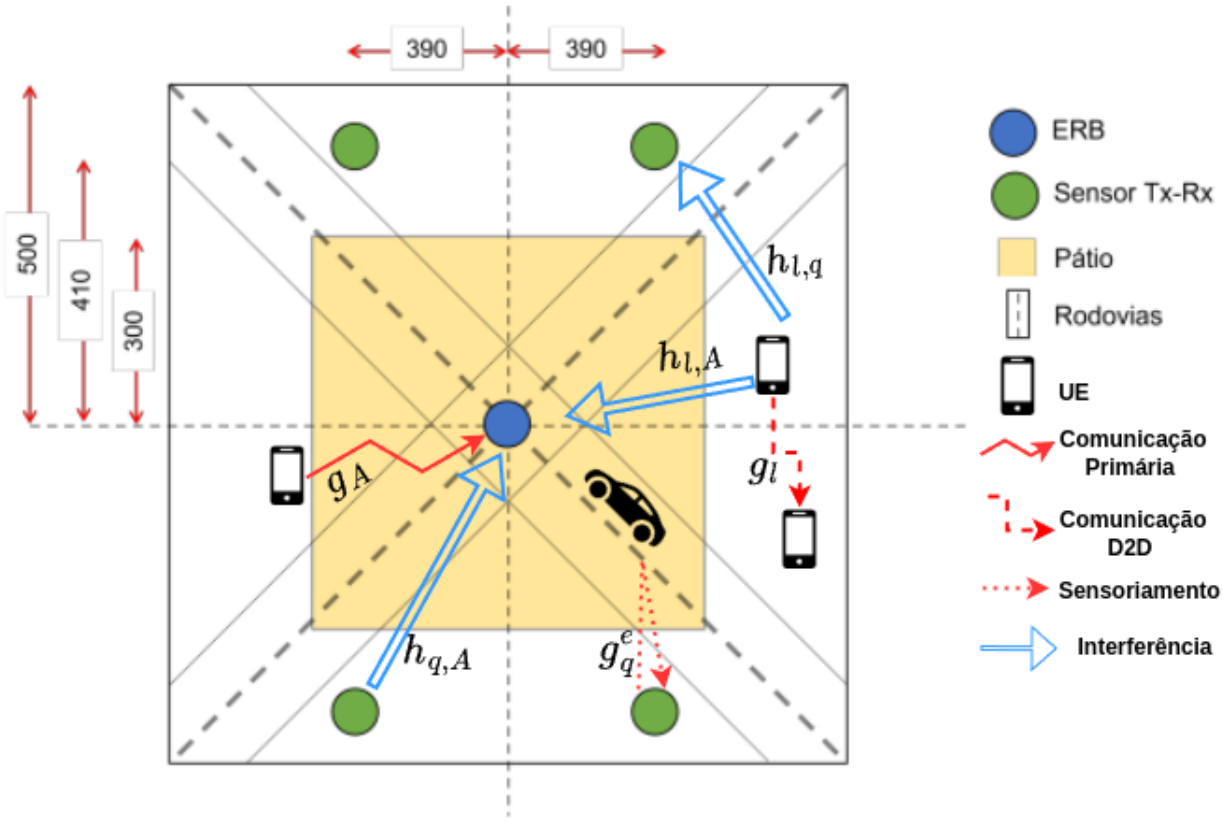

Figura 3.1 – Representação da célula modelada. Fonte: (Cardoso, 2024), com adaptações.

da l-ésima comunicação D2D e a ERB, entre o primeiro e o receptor do sensor q, e entre o transmissor do sensor q e a ERB; e representam, respectivamente, os níveis de interferência que a comunicação D2D gera na comunicação primária e no sensor, e a que o sensor gera na comunicação primária.

O modelo de mobilidade dos UEs segue o movimento browniano, enquanto os veículos, alvos do sensoreamento, realizam movimentos retilíneos e uniformes ao longo das vias. Caso algum elemento chegue à borda da célula a trajetória é refletida, assim como os veículos que alcancem o centro.

3.2.1 Canal

O modelo simula um fluxo contínuo de transmissão de dados entre os dispositivos, onde os pacotes são enfileirados em uma fila infinita e enviados em sequência. Os sensores são ativados quando o alvo entra na célula e são desligados quando o alvo sai da região de aplicação.

As comunicações, primária e D2D, e os sensores utilizam o mesmo canal. O sistema modela um único *resouce block* (RB), que pode ser compartilhado, a depender do cenário analisado, por duas ou três dessas entidades simultaneamente. Os sinais recebidos pela j-ésima comunicação e pelo q-ésimo sensor, quando presente no sistema, são definidos pelas

Equações 3.1 e 3.2.

$$y_j = \sqrt{p_j^t}\, g_j\, x_j + \sum_{i=1, i\neq j}^{J} \sqrt{p_i^t}\, h_{i,j}\, x_i\, I_j + \sum_{q=1}^{Q} \sqrt{p_q^t}\, h_{q,j}\, x_q\, I_q + \eta_j \tag{3.1}$$

$$y_q = \sqrt{p_q^t}\, g_q^e\, x_q + \sum_{j=1}^{J} \sqrt{p_j^t}\, h_{j,q}\, x_j\, I_j + \sum_{i=1, i\neq q}^{Q} \sqrt{p_i^t}\, h_{i,q}\, x_i + \eta_q, \tag{3.2}$$

onde $y_{[.]}$, $p_{[.]}^t$, $x_{[.]}$, $\eta_{[.]}$ e $I_{[.]}$ são o sinal recebido pela comunicação/sensor [.], sua potência de transmissão, o símbolo transmitido, o ruído recebido e o indicador de sua presença no cenário, e g_j refere-se a g_A ou g_l.

Em ambas equações, o primeiro termo, $\sqrt{p_{[.]}^t}\, g_{[.]}^*\, x_{[.]}$, representa o sinal enviado pelo transmissor para o receptor da comunicação/sensor [.]. O segundo, $\sum_*^J \sqrt{p_*^t}\, h_{*,[.]}\, x_*\, I_j$, representa a interferência por outras comunicações. O terceiro, $\sum_*^Q \sqrt{p_*^t}\, h_{*,[.]}\, x_*\, I_q$, representa a interferência dos sensores. Finalmente, o quarto, $\eta_{[.]}$, é o ruído ruído gaussiano branco aditivo (AWGN) no canal.

As relações sinal ruído mais interferência (SNIR) da comunicação j e do sensor q são definidas nas Equações 3.3 e 3.4.

$$\zeta_j = \frac{p_j^t\, |g_j|^2}{\sum_{i=1, i\neq j}^{J} p_i^t\, |h_{i,j}|^2 + \sum_{q=1}^{Q} p_q^t\, |h_{q,j}|^2 + \eta_j^2} \tag{3.3}$$

$$\zeta_q = \frac{p_q^t\, |g_q^e|^2}{\sum_{j=1}^{J} p_j^t\, |h_{j,q}|^2 + \sum_{i=1, i\neq q}^{Q} p_i^t\, |h_{i,q}|^2 + \eta_q^2}, \tag{3.4}$$

onde $\zeta_{[.]}$ é a SNIR do canal da comunicação/sensor [.].

O modelo de *path loss*, calculado na Equação 3.5 (Cardoso, 2024 apud MEI, 2018), foi baseado na especificação 3GPP TR 38.901 (Zhu *et al.*, 2021), que define modelos de propagação para cenários do 5G, especialmente na faixa de frequência de ondas milimétricas (mmWave), como a banda de 28 GHz, usada neste trabalho. O modelo de propagação escolhido foi projetado para cenários de quadrado aberto com linha de visada (LOS).

$$PL[dB] = 32{,}4 + 18{,}5\, log_{10}(d) + 20\, log_{10}(f_c), \tag{3.5}$$

onde d é a distância, em metros, entre o transmissor e o receptor e f_c é a frequência da portadora, em GHz.

Os ganhos de propagação, em dB, do canal direto, G, e do canal com reflexão em alvo, G^e, são definidos nas Equações 3.6 e 3.7.

$$G = G_t + G_r - (PL + SF) \tag{3.6}$$

$$G^e = G_t + G_r + \sigma_{RCS} - (PL_{T-A} + PL_{A-R} + SF), \tag{3.7}$$

onde G_t e G_r são os ganhos das antenas de transmissão do transmissor e de recepção do receptor, σ_{RCS} é a seção transversal do radar (RCS) dos alvos de sensoriamento, PL_{T-A} e PL_{A-R} são o *path loss* entre o transmissor do sensor e o alvo e entre o alvo e o receptor do sensor, e SF (*Shadow Fading Loss*) é a perda de desvanecimento devido ao sombreamento, que segue distribuição log-normal $\mathcal{N}(0, 4.2)$ (Cardoso, 2024 apud MEI, 2018) (Cardoso, 2024 apud Rappaport, 2024).

3.2.2 Requisitos

Para atingir os objetivos propostos para a comunicação primária e sensoriamento, devem ser atendidos requisitos mínimos. O primeiro, descrito na Equação 3.8 , é garantir que a comunicação primária tenha um canal com eficiência espectral ψ_a maior ou igual a um valor mínimo ψ_{min}, o que garante que a taxa de transmissão requisitada está sendo entregue (Cardoso, 2024 apud Shi *et al.*, 2020) (Cardoso, 2024 apud Fang *et al.*, 2022b). O segundo requisito é garantir, para o sensoriamento, que a probabilidade de detecção do alvo P_q^d, calculada na Equação 3.9 a partir do *generalized likelihood ratio test* (GLRT) (Cardoso, 2024 apud Li; Liu; Lei, 2023), seja maior que um limiar φ_{min} (Cardoso, 2024 apud Shi *et al.*, 2019) (Cardoso, 2024 apud Deligiannis *et al.*, 2017).

$$\psi_a = log_2(1 + \zeta_a) \geq \psi_{min} \tag{3.8}$$

$$P_q^d = \left(1 + \frac{\lambda}{1-\lambda}\frac{1}{1+\omega\zeta_q}\right)^{1-\omega} \geq \varphi_{min}, \tag{3.9}$$

onde λ é o limiar de detecção e w é o número de pulsos recebido por cada sensor durante o período de tempo no qual o alvo está sendo iluminado pela onda emitida pelo transmissor do sensor.

As comunicações D2D, por sua vez, não têm requisitos mínimos, uma vez que o objetivo é maximizar sua eficiência espectral de forma secundária, dando prioridade à comunicação primária e aos sensores.

3.3 Alocação de potência

A alocação é feita a partir do estado do canal. Para o caso centralizado, é assumido que a ERB conhece toda a informação do estado do canal (CSI). A cada intervalo de tempo pré-definido (*timestep*), o sistema amostra o ganho de propagação do canal direto entre o transmissor e o receptor de cada comunicação e o ganho de propagação do canal com reflexão de cada sensor. Além disso, a partir dessa técnica obtém-se o ganho de propagação dos canais interferentes das comunicações e dos sensores.

Por outro lado, no FL a alocação é feita considerando que a CSI de cada comunicação/-sensor é conhecida apenas pelos participantes, reduzindo a complexidade de implementação, detalhada no parágrafo acima, para obtenção das informações.

A decisão de alocação de potência é dinâmica, sendo executada a cada *timestep*, ou seja, atualizada de acordo com as mudanças do sistema, adaptando-se às condições de rede e demandas dos usuários, uma vez que os ganhos de propagação dos canais do sistema variam com o tempo, devido ao movimento dos elementos e de fatores aleatórios, como o *shadowing* e o desvanescimento de pequena escala (Cardoso, 2024).

Foram analisados três variações do problema de otimização conjunta da alocação de potência, representados em três cenários, detalhados a seguir, que alternam entre a presença de comunicação D2D, sensoriamento, ou ambos.

3.3.1 Cenário A: comunicação primária e sensoriamento

O primeiro cenário é constituído pela comunicação primária e os sensores, e tem por objetivo a maximização da eficiência espectral da comunicação primária ψ_a, com a restrição de que a probabilidade de detecção dos sensores P_q^d deve ser maior que a mínima φ_{min}, e as potências não devem exceder um valor máximo p_{max}, como descrito a seguir:

$$max_{p_A, p_q}(\psi_a) \tag{3.10}$$

$$t.q.\ P_q^d \geq \varphi_{min}, \quad \forall\, q \in Q \tag{3.11}$$

$$p_q \leq p_{max}, \quad \forall\, q \in Q \tag{3.12}$$

$$p_a \leq p_{max} \tag{3.13}$$

3.3.2 Cenário B: comunicações primária e D2D

O segundo cenário, composto pelas comunicações primária e D2D, tem o objetivo de maximizar a eficiência espectral da comunicação D2D ψ_l, enquanto mantém a eficiência espectral da comunicação primária acima da mínima ψ_{min}, e as potências abaixo da máxima, como modelado abaixo:

$$max_{p_j}(\Sigma_{l=1}^{L}\psi_l) \tag{3.14}$$

$$t.q.\ \psi_a \geq \psi_{min}, \quad \forall\, a \in A \tag{3.15}$$

$$p_j \leq p_{max}, \quad \forall\, j \in J \tag{3.16}$$

3.3.3 Cenário C: comunicações primária e D2D, e sensoriamento

O último cenário, mais complexo, é constituído pelas comunicações primária e D2D, e pelos sensores. O objetivo é maximizar a eficiência espectral das comunicações D2D,

respeitando as duas restrições, além do limite de potência: eficiência espectral mínima para comunicação primária e probabilidade mínima de detecção dos sensores. A modelagem matemática é mostrada a seguir:

$$max_{p_j,p_q}(\Sigma_{l=1}^{L}\psi_l) \tag{3.17}$$

$$t.q.\ \psi_a \geq \psi_{min}, \quad \forall\, a \in A \tag{3.18}$$

$$P_q^d \geq \varphi_{min}, \quad \forall\, q \in Q \tag{3.19}$$

$$p_j \leq p_{max}, \quad \forall\, j \in J \tag{3.20}$$

$$p_q \leq p_{max}, \quad \forall\, q \in Q \tag{3.21}$$

3.4 Aprendizado por reforço profundo federado

Todos os cenários apresentados acima representam problemas não convexos (Qin *et al.*, 2023) (Noman *et al.*, 2024) (Yang *et al.*, 2024a) e não lineares (Cardoso, 2024 apud Powell, 2022), o que torna difícil a solução por algoritmos tradicionais (Chen *et al.*, 2024) (Liu *et al.*, 2024).

Para lidar com o ambiente de alta complexidade, e levando em consideração a importância da privacidade dos dados dos usuários, este trabalho propõe uma solução baseada em aprendizado por reforço profundo aplicado com o paradigma de aprendizado federado do tipo horizontal, e compara com a forma centralizada, proposta em (Cardoso, 2024).

3.4.1 Estados

A representação do ambiente é adaptada para cada cenário e cada tipo de dispositivo que realiza o treinamento, uma vez que eles treinam apenas com dados locais, não acessando dados de outros dispositivos.

Para o cenário A, onde apenas a comunicação primária e os sensores estão presentes, o estado da primeira é descrito pelos ganhos de propagação do seu canal, g_a, e do canal entre o transmissor de cada sensor e a ERB, $h_{q,A}$, que são utilizados pelo algoritmo para compreender a situação da comunicação primária e o nível de interferência que cada sensor causa na mesma, respectivamente. Já o estado de cada sensor é representado pelo ganho de propagação do seu canal, g_q^e.

No cenário B, constituído pelas comunicações primária e D2D, o estado da primeira é análogo ao cenário anterior, com a diferença que a interferência agora é das comunicações D2D, representada pelos ganhos de propagação do canal entre cada transmissor D2D e a ERB, $h_{l,A}$. O estado de cada D2D, por sua vez, é a interferência total sentida pela ERB, representada

pela soma dos ganhos de todos os canais de interferência, $\sum_{l=1}^{L} h_{l,A}$, sendo esta informação obtida a partir de um *broadcasting* no ínicio de cada rodada de treinamento.

Finalmente, no cenário C, que é composto pelas duas comunicações e sensores, o estado da comunicação primária é descrito apenas pelo ganho de propagação do seu canal, enquanto o estado de cada D2D continua sendo o somatório dos ganhos dos canais de interferência com a ERB, incluindo dos sensores, $\sum_{l=1}^{L} h_{l,A} + \sum_{q=1}^{Q} h_{q,A}$. Por fim, o estado de cada sensor é o mesmo do cenário A.

3.4.2 Ação

A ação, a cada *timestep*, é um vetor que representa a alocação de potência para cada comunicação ou sensor. A dimensão do vetor é 10, pois é o número máximo de potências a serem alocadas. Quando há menos comunicações e/ou sensores do que o máximo, os elementos correspondentes no vetor não são utilizados.

3.4.3 Recompensa

Similarmente aos estados, as recompensas também variam de acordo com o cenário e o tipo de dispositivo, e foram adaptadas de forma empírica a partir de uma equação base, que premia o comportamento esperado, enquanto penaliza o comportamento indesejado.

A recompensa da comunicação primária, r_a, mostrada na Equação 3.22, tem um fator positivo caso a SNIR seja maior que a mínima, ou seja, não esteja em *outage*, seguindo o requisito de eficiência espectral (Cardoso, 2024), mas desconta um valor proporcional à interferência total recebida, $h_{total,A}$. A recompensa dos sensores, descrita na Equação 3.23, segue o mesmo princípio de proteção contra *outage*, mas sem descontar a interferência. Já a recompensa para as comunicações D2D, calculada na Equação 3.24, é igual a sua SNIR descontando a interferência total.

$$r_a = \hat{r}_a - 10h_{total,A}, \tag{3.22}$$

onde

$$\hat{r}_a = \begin{cases} 5 - 0.1\zeta_a & se \;\; \zeta_a > \zeta_{min}^{A} \\ -\zeta_{min}^{A} + 0.5\zeta_a & caso\ contrário \end{cases}$$

$$r_q = \begin{cases} 2.5 - 0.05\zeta_q & se \;\; \zeta_q > \zeta_{min}^{Q} \\ -0.05\zeta_{min}^{Q} + 0.5\zeta_q & caso\ contrário \end{cases} \tag{3.23}$$

$$r_l = \zeta_l - 10h_{total,A} \tag{3.24}$$

Para o caso centralizado, a recompensa r_C é a soma ponderada das recompensas de todos os dispositivos individuais, como mostrado na Equação 3.25.

$$r_C = w_1 r_a + w_2 \sum_{q=1}^{Q} r_q + w_3 \sum_{l=1}^{L} r_l, \tag{3.25}$$

onde w_1, w_2 e w_3 são os pesos que ponderam cada fator da função.

3.4.4 Agregação

A agregação do modelo global é feita a partir de uma média ponderada dos parâmetros de rede dos modelos locais, como elaborado nas Equações 3.26, 3.27 e 3.28.

$$\theta_g = \frac{w_a\, \theta_a + w_l\, \hat{\theta}_l + w_q\, \hat{\theta_q}}{w_a + w_l + w_q} \tag{3.26}$$

$$\hat{\theta}_l = \frac{\sum_{l=1}^{L} \theta_l}{L} \tag{3.27}$$

$$\hat{\theta_q} = \frac{\sum_{q=1}^{Q} \theta_q}{Q}, \tag{3.28}$$

onde, θ_g, θ_a, θ_l e θ_q são os parâmetros de modelo global, da comunicação primária, D2D e do sensor, e w_a, w_l e w_q são os respectivos pesos, definidos de forma empírica.

3.4.5 Simulação

Os parâmetros definidos para a simulação do ambiente são mostrados na Tabela 3.1.

Com o objetivo de que o algoritmo se adapte a diversas situações do sistema de comunicação, a quantidade de comunicações D2D e de sensores presentes nas simulações variam de acordo com distribuições de Poisson, com taxas de ocorrência iguais a 4 e 2, respectivamente. A quantidade de comunicações primárias é constante e igual a 1.

A partir da definição do número de comunicações, há a inicialização dos dispositivos no ambiente, da seguinte forma: a BS e os sensores são inicializados nas suas posições, detalhadas na Figura 3.1; a posição do UE é inicializada em algum lugar no pátio, com probabilidade uniformemente distribuída, e com velocidades que seguem uma distribuição uniforme $\mathcal{U}_{[-20,20]}$ km/h, para cada eixo do plano cartesiano; as posições dos transmissores D2D seguem a mesma lógica, com velocidades seguindo uma distribuição uniforme $\mathcal{U}_{[-5,5]}$ km/h, enquanto os receptores D2D são inicializados próximos aos respectivos transmissores, com distância que segue distribuição $\mathcal{U}_{[-20,20]}$ em cada eixo do plano, e suas velocidades seguem a mesma distribuição dos transmissores; por último, os alvos tem igual probabilidade de serem inicializados em qualquer parte das rodovias fora do pátio, com velocidades constantes iguais a 40 km/h.

Foi definido que nenhum D2D pode se aproximar a um raio menor que 100 metros da BS, para evitar um alto nível de interferência das comunicações primárias em *uplink*.

Tabela 3.1 – Parâmetros para a implementação do ambiente. Fonte: (Cardoso, 2024), com adaptações.

Parâmetro	Valor
ψ_{min}	2,6 bps/Hz
ζ_{min}^{A}	5 dB
φ_{min}	0,99
w	512 pulsos/s
λ	0,001
ζ_{min}^{Q}	-10 dB
w_1	2
w_2	1,5
w_3	0,1
f_c	28 GHz
G_t^{sensor}	32
G_t^{D2D}	0
G_r^{BS}	32
G_r^{sensor}	32
G_r^{D2D}	0
SF	$\mathcal{N}(0;4,2)$
σ_{RCS}	1
p_{max}	50 dBm
w_a	12
w_l	1
w_q	5

As simulações foram feitas com *timestep* de 1 segundo, então cada comunicação e sensor têm suas potências realocadas nesse intervalo.

3.4.6 Configuração dos algoritmos

A estrutura comum base das redes para os agente dos algoritmos utilizados, PPO e REINFORCE, foi feita da seguinte forma: a camada inicial é linear, transformando a dimensão do espaço de estados em um espaço de 64 unidades; normalização de camada e ativação *Rectified Linear Unit* (ReLU) para estabilizar o aprendizado e introduzir não linearidade; depois, há a expansão para 128 unidades, e subsequente redução para 64, novamente seguida de normalização e ativação; por fim, a camada final é linear, que mapeia as unidades da camada anterior para o espaço de ações (Cardoso, 2024).

Para o algoritmo PPO, também foi implementada uma rede que atua como crítico, com arquitetura similar, mas com a camada de saída de dimensão igual a 1, pois é utilizada apenas para a estimação da função valor (V).

Os principais parâmetros utilizados para o treinamento dos dois algoritmos estão descritos na Tabela 3.2.

Tabela 3.2 – Parametrização dos Algoritmos PPO e REINFORCE. Fonte: (Cardoso, 2024), adaptado.

Parâmetro	PPO	REINFORCE
Episódios de treinamento	2000	2000
Fator de desconto do retorno (γ)	0,9	0,99
Fator de suavização da atualização (τ)	1	-
Desvio padrão inicial do ruído de exploração (σ_0)	20	15
Coeficiente de adaptação do ruído de exploração	1,02	1,05
Intervalo para adaptação do ruído de exploração (T_{adapt})	8	5
Desvio padrão mínimo do ruído de exploração	0,001	0,001
Otimizadores das redes neurais do crítico	Adam	Adam (Diederik, 2014)
Taxa de aprendizagem do crítico	0,001	-
Epocas de iteração para treinamento das redes (K)	50	-
Limitantes da função de perda do agente (ε)	0,2	-

3.5 Conclusão

Neste capítulo, foram definidos os elementos essenciais para a implementação do modelo proposto, incluindo a caracterização do sistema de comunicações, os desafios da alocação de potência e a formulação dos cenários estudados. A modelagem estabelece as condições necessárias para a aplicação do aprendizado por reforço profundo federado, evidenciando a complexidade do problema e a importância de abordagens inteligentes para sua resolução. A estrutura descrita aqui servirá como base para a análise dos resultados nos capítulos seguintes, possibilitando a avaliação do impacto da estratégia proposta na otimização do desempenho da rede JCAS, quando comparada com a solução centralizada análoga.

4 Análise e discussão de resultados

4.1 Introdução

Este capítulo apresenta os resultados de treinamento dos modelos nos cenários estudados, descritos no Capítulo 3. São feitas comparações entre os desempenhos dos algoritmos, PPO e REINFORCE, e paradigmas de treinamento, federado e centralizado, para cada tipo de situação.

As métricas essenciais, i.e. presentes em todos os cenários, utilizadas para a comparação são a probabilidade de *outage*, que a partir daqui será referida apenas como *outage*, e SNIR das comunicações primárias. As demais dependem do cenário analisado, são elas: *outage* e SNIR dos sensores, e SNIR das comunicações D2D. A definição usada neste trabalho para *outage* de uma comunicação é o percentual de vezes nas quais o SNIR da comunicação esteve abaixo do mínimo estabelecido para a mesma. Os resultados de SNIR, por sua vez, representam as médias aritméticas da SNIR de todas as comunicações de mesmo tipo.

Foram treinadas, por cenário, 5 instâncias de cada algoritmo em cada paradigma. Os resultados apresentados nas Tabelas 4.1, 4.2 e 4.3 são as médias aritméticas dos últimos 200 episódios desses treinamentos para os cenários A, B e C, respectivamente.

4.2 Cenário A: comunicação primária e sensoriamento

O primeiro cenário é constituído apenas pela comunicação primária e sensores. Os resultados do treinamento dessa configuração são apresentados na Tabela 4.1, a seguir.

Tabela 4.1 – Resultado de treinamento para o cenário A.

Algoritmo	PPO		REINFORCE	
Variável	Centralizado	Federado	Centralizado	Federado
Outage comunicação primária (%)	3,77	9,22	42,29	47,46
Outage sensor (%)	0,33	5,12	0	0
SNIR comunicação primária (dB)	14,71	27,45	5,40	4,67
SNIR Sensores (dB)	1,07	2,49	9,39	9,26

O algoritmo PPO mostrou desempenho superior ao REINFORCE no objetivo principal de maximizar a eficiência espectral da comunicação primária, reduzindo o *outage* de 42,29% para 3,77% no caso centralizado e 47,46% para 9,22%, no federado, além de aumento de 9,31 dB e 22,78 dB, respectivamente, no SNIR da mesma. Por outro lado, o REINFORCE apresentou melhor resultado no sensoriamento, zerando o outage dos sensores, com reduções de 0,33% e 5, 12%, e aumentando o SNIR em 8,32% e 6,77%, nos modelos centralizado e

federado, respectivamente. Em relação ao tipo de treino, o federado apresentou desempenho geral um pouco pior, com aumentos de 5,45% e 5,17% no *outage* da comunicação primária no PPO e REINFORCE, respectivamente, além de *outage* do sensor 4,79% maior para o PPO, enquanto no REINFORCE se manteve zerado. Por outro lado, o PPO federado apresentou desempenho melhor nas SNIRs quando comparado com o centralizado, com aumentos de 12,74 dB e 1,42 dB, para comunicação primária e sensoriamento, respectivamente, enquanto no REINFORCE houve reduções de 0,73 dB e 0,13 dB, 1,38%.

4.3 Cenário B: comunicações primária e D2D

O segundo cenário é constituído pelas comunicações primária e D2D. Os resultados do treinamento são apresentados na Tabela 4.2, a seguir.

Tabela 4.2 – Resultado de treinamento para o cenário B.

Algoritmo	PPO		REINFORCE	
Variável	Centralizado	Federado	Centralizado	Federado
Outage comunicação primária (%)	7,04	8,41	0,2	0.155
SNIR comunicação primária (dB)	12,54	12,11	27,75	26.24
SNIR D2D (dB)	4,16	0,09	-10,16	-8,84

Em contraste com o cenário A, o REINFORCE demonstrou resultados melhores para o objetivo principal, com redução nos *outages* de 7,02% no centralizado e 8,25% no federado, em comparação ao PPO. Além disso, também obteve melhor desempenho quanto à SNIR da comunicação primária, com aumentos de 15,21 dB e 14,13 dB. Já o PPO foi melhor no desempenho da comunicação D2D, com aumentos de 14,32 dB e 8,93 dB no SNIR.

Por outro lado, as diferenças entre centralizado e federado diminuíram nesse cenário, com o *outage* e SNIR da comunicação primária do federado sendo apenas 0,71% maior e 0,43 dB menor no PPO, e 0,045% e 1,51 dB menores no REINFORCE. Já para a SNIR da comunicação D2D, o PPO e REINFORCE federados obtiveram redução de 4,07 dB e aumento de 1,32 dB, respectivamente.

4.4 Cenário C: comunicações primária e D2D, e sensoriamento

A Tabela 4.3 mostra os resultados de treinamento no cenário completo, composto pelas comunicações primária e D2D, e sensoriamento.

Similarmente ao cenário A, o algoritmo PPO mostrou melhor desempenho no objetivo principal, com uma reduções de 52,52% , centralizado, e 42,08%, federado, no *outage* da comunicação primária , e aumentos de 9,43 dB e 7,52 dB na SNIR da mesma, enquanto

Tabela 4.3 – Resultado de treinamento para o cenário C.

Algoritmo	PPO		REINFORCE	
Variável	Centralizado	Federado	Centralizado	Federado
Outage comm primária (%)	4,13	11,27	56,65	53,35
Outage sensor (%)	0,92	0,14	0	0,04
SNIR comm primária (dB)	13,96	12,17	4,53	4,64
SNIR D2D (dB)	-20,72	-14,24	-12,43	-12,47
SNIR Sensores (dB)	0,97	-3,59	9,77	9,06

o REINFORCE apresentou diminuição de 0,92% e 0,10% no *outage* do sensor, e aumentos de 8,8 dB e 12, 65 dB na SNIR. Além disso, o REINFORCE também melhorou a SNIR da comunicação D2D, com aumentos de 8,29 dB e 1,77 dB.

Ademais, o paradigma federado no algoritmo PPO apresentou um aumento de 7,14% e diminuição de 0,78% nos *outages* da comunicação primária e dos sensores, enquanto no algoritmo REINFORCE houve uma redução de 3,3% e aumento de 0,04 %, respectivamente. Em relação às SNIRs das comunicações primária e D2D, e dos sensores, o PPO federado teve redução de 1,79 dB, aumento de 6,48 dB e redução de 4,56 dB, respectivamente, enquanto no REINFORCE teve aumento de 0,11 dB, e reduções de 0,04 dB e 0,71 dB.

Como esperado, os resultados demonstram, no geral, uma piora generalizada, quando comparados com os outros cenários, devido à maior complexidade do problema.

4.5 Conclusão

A partir dos resultados apresentados nesta seção, é possível observar que o algoritmo PPO mostrou maior estabilidade no objetivo primário de proteção da comunicação primária, obtendo resultados com menor variação entre os diferentes cenários , e melhor desempenho significativo em dois dos cenários analisados, enquanto o REINFORCE obteve um melhor desempenho menos expressivo na proteção dos sensores.

Quanto ao paradigma federado, os resultados demonstraram, no geral, pequena piora no desempenho do PPO em relação ao objetivo principal, evidenciado pelo *outage* das comunicações primárias, o que já era esperado, uma vez que o fato do modelo global não aprender diretamente dos dados no FL dificulta o mapeamento entre entradas e saídas. Entretanto, a diferença observada entre os paradigmas federado e centralizado não são muito expressivos, o que torna o FL viável, tendo em vista as vantagens na proteção dos dados dos usuários.

5 Conclusão

Neste trabalho, foi investigada a alocação de potência em redes JCAS utilizando aprendizado por reforço profundo federado. Foram considerados diferentes cenários e comparados os desempenhos dos algoritmos PPO e REINFORCE em paradigmas centralizado e federado.

Os resultados demonstraram que o algoritmo PPO apresentou maior estabilidade e desempenho na proteção da comunicação primária, enquanto o REINFORCE obteve um desempenho melhor na proteção dos sensores. Além disso, a comparação entre os paradigmas resultou em uma leve diminuição no desempenho, que não é expressivo para invalidar o uso do FL, que proporciona benefícios de maior privacidade e segurança para os dados dos usuários, além de não precisar de conhecimento total da CSI, o que reduz a complexidade do sistema.

Com esses resultados, este estudo buscou contribuir com as discussões sobre o emprego de técnicas de DRL e FL para a evolução das estratégias de alocação de potência em redes JCAS com múltiplas comunicações, visando a otimização dos desempenhos das comunicações e sensoriamento e a privacidade de dados dos usuários.

Como trabalhos futuros, sugere-se:

- Investigação de outros métodos de agregação no estado da arte para o FL, visando otimização do desempenho das comunicações e sensores;
- Análise comparativa de métricas relacionadas a eficiência energética e *throughput*, entre aprendizados federado e centralizado.

Referências

AL-QURAAN, M.; MOHJAZI, L.; BARIAH, L.; CENTENO, A.; ZOHA, A.; MUHAIDAT, S.; DEB-BAH, M.; IMRAN, M. A. Edge-native intelligence for 6g communications driven by federated learning: A survey of trends and challenges. arxiv 2021. **arXiv preprint arXiv:2111.07392**, 2021. Citado na p. 28.

ALENEZI, S.; LUO, C.; MIN, G. Energy-efficient d2d communications based on centralised reinforcement learning techniques. *In*: IEEE. **2021 IEEE 24th International Conference on Computational Science and Engineering (CSE)**. [*S.l.*], 2021. p. 57–63. Citado nas pp. 16 e 19.

ALSALEH, S.; MENAI, M. E. B.; AL-AHMADI, S. Federated learning–based model to lightweight idss for heterogeneous iot networks: State-of-the-art, challenges and future directions. **IEEE Access**, IEEE, 2024. Citado na p. 27.

BHATIA, L.; SAMET, S. Decentralized federated learning: A comprehensive survey and a new blockchain-based data evaluation scheme. *In*: IEEE. **2022 Fourth International Conference on Blockchain Computing and Applications (BCCA)**. [*S.l.*], 2022. p. 289–296. Citado na p. 27.

BUSONIU, L.; BABUSKA, R.; SCHUTTER, B. D. A comprehensive survey of multiagent reinforcement learning. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, IEEE, v. 38, n. 2, p. 156–172, 2008. Citado na p. 24.

CARDOSO, G. P. d. F. Deep reinforcement learning e hiper-heurística aplicados à alocação de recursos em sistemas de comunicações 6g com comunicações d2d e sensoreamento. **Disseratação de Mestrado em Engenharia Elétrica, UnB**, 2024. Citado nas pp. 6, 7, 12, 13, 16, 17, 20, 31, 32, 33, 34, 35, 36, 37, 39 e 40.

CHEN, Y.; YANG, H.; OU, X.; JIANG, Y.; XIONG, Z. Anti-jamming resource allocation for integrated sensing and communications based on game-guided reinforcement learning. **IEEE Wireless Communications Letters**, IEEE, 2024. Citado nas pp. 14, 18 e 36.

CHENG, K.; ZHAO, Y.; WANG, B.; LIANG, C. Dynamic environment-adaptive uav-assisted integrated sensing and communication. *In*: IEEE. **2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)**. [*S.l.*], 2024. p. 1–5. Citado nas pp. 14 e 18.

DAS, S. K.; CHAMPAGNE, B.; PSAROMILIGKOS, I.; CAI, Y. A survey on federated learning for reconfigurable intelligent metasurfaces-aided wireless networks. **IEEE Open Journal of the Communications Society**, IEEE, 2024. Citado na p. 27.

DELIGIANNIS, A.; PANOUI, A.; LAMBOTHARAN, S.; CHAMBERS, J. A. Game-theoretic power allocation and the nash equilibrium analysis for a multistatic mimo radar network. **IEEE Transactions on Signal Processing**, IEEE, v. 65, n. 24, p. 6397–6408, 2017. Citado na p. 34.

DIEDERIK, P. K. Adam: A method for stochastic optimization. **(No Title)**, 2014. Citado na p. 40.

ELDAR, Y. C.; GOLDSMITH, A.; GÜNDÜZ, D.; POOR, H. V. **Machine learning and wireless communications**. [*S.l.*]: Cambridge University Press, 2022. Citado na p. 22.

ENGELBRECHT, A. P. **Computational intelligence: an introduction**. [*S.l.*]: John Wiley & Sons, 2007. Citado na p. 23.

FANG, X.; FENG, W.; CHEN, Y.; GE, N.; ZHANG, Y. Joint communication and sensing toward 6g: Models and potential of using mimo. **IEEE Internet of Things Journal**, IEEE, v. 10, n. 5, p. 4093–4116, 2022. Citado nas pp. 12 e 29.

FANG, X.; FENG, W.; CHEN, Y.; MA, D.; GE, N.; LIU, Y.; FENG, Z. Radio map-based spectrum sharing for joint communication and sensing networks. *In*: IEEE. **2022 IEEE/CIC International Conference on Communications in China (ICCC)**. [*S.l.*], 2022. p. 238–243. Citado na p. 34.

GONZÁLEZ-PRELCIC, N.; KESKIN, M. F.; KALTIOKALLIO, O.; VALKAMA, M.; DARDARI, D.; SHEN, X.; SHEN, Y.; BAYRAKTAR, M.; WYMEERSCH, H. The integrated sensing and communication revolution for 6g: Vision, techniques, and applications. **Proceedings of the IEEE**, IEEE, 2024. Citado nas pp. 12, 29 e 30.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [*S.l.*]: MIT Press, 2016. http://www.deeplearningbook.org. Citado na p. 26.

GUO, Q.; TANG, F.; KATO, N. Federated reinforcement learning-based resource allocation in d2d-enabled 6g. **IEEE Network**, IEEE, v. 37, n. 5, p. 89–95, 2022. Citado nas pp. 16 e 20.

HUANG, Y.; FANG, Y.; LI, X.; XU, J. Coordinated power control for network integrated sensing and communication. **IEEE Transactions on Vehicular Technology**, IEEE, v. 71, n. 12, p. 13361–13365, 2022. Citado nas pp. 13 e 18.

JENO, G. **Federated Learning with Python: Design and implement a federated learning system and develop applications using existing frameworks**. [*S.l.*]: Packt Publishing Ltd, 2022. Citado nas pp. 6, 27 e 28.

KIM, H.; HWANG, M.; JEE, J.; PARK, J.; PARK, H. 3d state transition modeling and power allocation for uav-aided isac system. *In*: IEEE. **2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)**. [*S.l.*], 2023. p. 1–6. Citado nas pp. 14 e 18.

LI, M.; LIU, W.; LEI, J. A review on orthogonal time–frequency space modulation: State-of-art, hotspots and challenges. **Computer Networks**, Elsevier, v. 224, p. 109597, 2023. Citado na p. 34.

LI, Q.; WEN, Z.; WU, Z.; HU, S.; WANG, N.; LI, Y.; LIU, X.; HE, B. A survey on federated learning systems: Vision, hype and reality for data privacy and protection. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 35, n. 4, p. 3347–3366, 2021. Citado na p. 27.

LI, T.; ZHU, K.; LUONG, N. C.; NIYATO, D.; WU, Q.; ZHANG, Y.; CHEN, B. Applications of multi-agent reinforcement learning in future internet: A comprehensive survey. **IEEE Communications Surveys & Tutorials**, IEEE, v. 24, n. 2, p. 1240–1279, 2022. Citado na p. 24.

LIM, W. Y. B.; LUONG, N. C.; HOANG, D. T.; JIAO, Y.; LIANG, Y.-C.; YANG, Q.; NIYATO, D.; MIAO, C. Federated learning in mobile edge networks: A comprehensive survey. **IEEE communications surveys & tutorials**, IEEE, v. 22, n. 3, p. 2031–2063, 2020. Citado nas pp. 12, 25, 26 e 27.

LIU, C.; XIA, M.; ZHAO, J.; LI, H.; GONG, Y. Optimal resource allocation for integrated sensing and communications in internet of vehicles: A deep reinforcement learning approach. **IEEE Transactions on Vehicular Technology**, IEEE, 2024. Citado nas pp. 15, 18 e 36.

LIU, M.; YANG, M.; WEI, F.; LI, H.; ZHANG, Z.; NALLANATHAN, A.; HANZO, L. A non-orthogonal uplink/downlink iot solution for next-generation isac systems. **IEEE Internet of Things Journal**, IEEE, 2023. Citado nas pp. 13 e 18.

LIU, Y.; YU, F. R.; LI, X.; JI, H.; LEUNG, V. C. Blockchain and machine learning for communications and networking systems. **IEEE Communications Surveys & Tutorials**, v. 22, n. 2, p. 1392–1431, 2020. Citado nas pp. 6, 22, 23 e 24.

MEI, F. 28 ghz applications, path loss models and coverage for 5g. Politecnico di Milano, 2018. Citado nas pp. 33 e 34.

MUNIKOTI, S.; AGARWAL, D.; DAS, L.; HALAPPANAVAR, M.; NATARAJAN, B. Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications. **IEEE transactions on neural networks and learning systems**, IEEE, 2023. Citado na p. 26.

NIE, P.-y.; ZHANG, P.-a. A note on stackelberg games. *In*: IEEE. **2008 Chinese Control and Decision Conference**. [*S.l.*], 2008. p. 1201–1203. Citado na p. 14.

NITHYA, T.; KUMAR, V. N.; GAYATHRI, S.; DEEPA, S.; VARUN, C.; SUBRAMANIAN, R. S. A comprehensive survey of machine learning: Advancements, applications, and

challenges. *In*: IEEE. **2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)**. [*S.l.*], 2023. p. 354–361. Citado na p. 23.

NOMAN, H. M. F.; DIMYATI, K.; NOORDIN, K. A.; HANAFI, E.; ABDRABOU, A. Fedrl-d2d: Federated deep reinforcement learning-empowered resource allocation scheme for energy efficiency maximization in d2d-assisted 6g networks. **IEEE Access**, IEEE, 2024. Citado nas pp. 16, 19 e 36.

POWELL, W. B. Policy function approximations and policy search. Wiley Data and Cybersecurity, 2022. Citado na p. 36.

QIN, Y.; ZHANG, Z.; LI, X.; HUANGFU, W.; ZHANG, H. Deep reinforcement learning based resource allocation and trajectory planning in integrated sensing and communications uav network. **IEEE Transactions on Wireless Communications**, IEEE, v. 22, n. 11, p. 8158–8169, 2023. Citado nas pp. 14, 18 e 36.

RAPPAPORT, T. S. **Wireless communications: principles and practice**. [*S.l.*]: Cambridge University Press, 2024. Citado na p. 34.

SCHULMAN, J.; WOLSKI, F.; DHARIWAL, P.; RADFORD, A.; KLIMOV, O. Proximal policy optimization algorithms. **arXiv preprint arXiv:1707.06347**, 2017. Citado na p. 27.

SHI, C.; QIU, W.; WANG, F.; SALOUS, S.; ZHOU, J. Power control scheme for spectral coexisting multistatic radar and massive mimo communication systems under uncertainties: A robust stackelberg game model. **Digital Signal Processing**, Elsevier, v. 94, p. 146–155, 2019. Citado na p. 34.

SHI, C.; WANG, Y.; WANG, F.; SALOUS, S.; ZHOU, J. Joint optimization scheme for subcarrier selection and power allocation in multicarrier dual-function radar-communication system. **IEEE Systems Journal**, IEEE, v. 15, n. 1, p. 947–958, 2020. Citado na p. 34.

SU, Y.; LÜBKE, M.; FRANCHI, N. Coordinated multipoint jcas in 6g mobile networks. **IEEE Access**, IEEE, 2024. Citado na p. 31.

SURAKHI, O. M.; GARCÍA, A. M.; JAMOOS, M.; ALKHANAFSEH, M. Y. A comprehensive survey for machine learning and deep learning applications for detecting intrusion detection. *In*: IEEE. **2021 22nd International Arab Conference on Information Technology (ACIT)**. [*S.l.*], 2021. p. 1–13. Citado na p. 25.

TAYLOR, H. The cramer-rao estimation error lower bound computation for deterministic nonlinear systems. decision and control including the 17th symposium on adaptive processes. *In*: **1978 IEEE Conference on**. [*S.l.*: *s.n.*], 1978. v. 17, p. 1178–1181. Citado na p. 14.

UPRETY, A.; RAWAT, D. B. Reinforcement learning for iot security: A comprehensive survey. **IEEE Internet of Things Journal**, IEEE, v. 8, n. 11, p. 8693–8706, 2020. Citado na p. 24.

WANG, X.; WANG, S.; LIANG, X.; ZHAO, D.; HUANG, J.; XU, X.; DAI, B.; MIAO, Q. Deep reinforcement learning: A survey. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, v. 35, n. 4, p. 5064–5078, 2022. Citado na p. 26.

WANG, X.; WU, H.; XU, Y.; CAO, H.; KUMAR, N.; RODRIGUES, J. J. Resource allocation in multi-cell integrated sensing and communication systems: A drl approach. *In*: IEEE. **ICC 2023-IEEE International Conference on Communications**. [*S.l.*], 2023. p. 3210–3215. Citado nas pp. 12, 15 e 19.

WILLIAMS, R. J. Simple statistical gradient-following algorithms for connectionist reinforcement learning. **Machine learning**, Springer, v. 8, p. 229–256, 1992. Citado na p. 27.

XIE, J.; YU, F. R.; HUANG, T.; XIE, R.; LIU, J.; WANG, C.; LIU, Y. A survey of machine learning techniques applied to software defined networking (sdn): Research issues and challenges. **IEEE Communications Surveys & Tutorials**, IEEE, v. 21, n. 1, p. 393–430, 2018. Citado nas pp. 23 e 24.

YANG, H.; WANG, L.; FENG, Z.; WEI, Z.; PENG, J.; YUAN, X.; QUEK, T. Q.; ZHANG, P. Dynamic power allocation for integrated sensing and communication-enabled vehicular networks. **IEEE Transactions on Wireless Communications**, IEEE, 2024. Citado nas pp. 15, 19 e 36.

YANG, L.; WEI, Y.; FENG, Z.; ZHANG, Q.; HAN, Z. Deep reinforcement learning-based resource allocation for integrated sensing, communication, and computation in vehicular network. **IEEE Transactions on Wireless Communications**, IEEE, 2024. Citado nas pp. 16 e 19.

ZHANG, J. A.; RAHMAN, M. L.; WU, K.; HUANG, X.; GUO, Y. J.; CHEN, S.; YUAN, J. Enabling joint communication and radar sensing in mobile networks—a survey. **IEEE Communications Surveys & Tutorials**, IEEE, v. 24, n. 1, p. 306–345, 2021. Citado na p. 29.

ZHAO, C.; WU, G.; XIONG, W. Decentralized multiagent reinforcement learning-based cooperative perception with dual-functional radar-communication v2v links. *In*: IEEE. **2023 IEEE International Conference on Communications Workshops (ICC Workshops)**. [*S.l.*], 2023. p. 1100–1105. Citado nas pp. 15 e 19.

ZHU, Q.; WANG, C.-X.; HUA, B.; MAO, K.; JIANG, S.; YAO, M. 3gpp tr 38.901 channel model. *In*: **the wiley 5G Ref: the essential 5G reference online**. [*S.l.*]: Wiley Press Hoboken, NJ, USA, 2021. p. 1–35. Citado na p. 33.

ZUO, Y.; GUO, J.; GAO, N.; ZHU, Y.; JIN, S.; LI, X. A survey of blockchain and artificial intelligence for 6g wireless communications. **IEEE Communications Surveys & Tutorials**, IEEE, 2023. Citado na p. 24.


UnB