



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma Solução de Segurança Operacional para a Mobilidade Aérea Urbana com Aprendizado de Reforço Profundo

Cristiano Perez Garcia

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Orientador
Prof. Dr. Li Weigang

Brasília
2023

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

G216s Garcia, Cristiano Perez
Uma Solução de Segurança Operacional para a Mobilidade
Aérea Urbana com Aprendizado de Reforço Profundo / Cristiano
Perez Garcia; orientador Li Weigang. -- Brasília, 2023.
109 p.

Dissertação(Mestrado em Informática) -- Universidade de
Brasília, 2023.

1. Mobilidade aérea urbana. 2. Aprendizado de reforço
profundo. 3. Resolução de conflitos. 4. Gerenciamento de
Tráfego Aéreo. 5. . I. Weigang, Li, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Uma Solução de Segurança Operacional para a Mobilidade Aérea Urbana com Aprendizado de Reforço Profundo

Cristiano Perez Garcia

Dissertação apresentada como requisito parcial para
conclusão do Mestrado em Informática

Prof. Dr. Li Weigang (Orientador)
CIC/UnB

Prof. Dr. Marcelo Xavier Guterres Prof. Dr. Marcelo Antonio Marotta
Instituto Tecnológico de Aeronáutica CIC/UnB

Prof. Dr. Ricardo Pezzuol Jacobi
Coordenador do Programa de Pós-graduação em Informática

Brasília, 16 de junho de 2023

Dedicatória

Dedico este trabalho à minha família. Em especial à minha mãe Janete, que sempre acreditou nos meus sonhos e tem sido parte das conquistas até aqui. À esposa Maisa cujo suporte me permitiu seguir essa jornada de forma muito mais suave. E aos animais de estimação Sheldon, Jenny e Lisa que acompanharam o meu trabalho por muitas horas e apoiaram mesmo sem compreender.

Agradecimentos

Agradeço ao meu orientador, Prof. Dr. Li Weigang pela valorosa contribuição no trabalho com todo o seu conhecimento e experiência.

Às professoras e professores do CIC/UnB, em especial, ao professor Geraldo Pereira Rocha Filho, que acompanhou o desenvolvimento do trabalho e forneceu orientações significativas para o resultado final durante o exame de qualificação.

Ao colega do TransLab Lucas Borges Monteiro, com quem compartilhei parte da caminhada e que deu sugestões bastante importantes para o aprimoramento do trabalho.

À Força Aérea Brasileira (FAB), na qual ganhei a experiência profissional que permitiu a escolha do tema. Na FAB, agradeço aos Chefes que acreditaram que a realização do Mestrado seria importante, em especial, ao Coronel Aviador Anderson Belchior Zuchetto de Castro e ao Tenente-Coronel Especialista em Tráfego Aéreo Miguel Antunes da Costa Júnior.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

Aeronaves com propulsão elétrica e com capacidade de efetuar pousos e decolagens na vertical, também conhecidos como eVTOL estão em desenvolvimento por uma série de fabricantes e tem o potencial de revolucionar a mobilidade aérea urbana nos próximos anos. A adoção tende a ser gradual, porém uma vez que um certo nível de maturidade desse tipo de transporte seja atingido, a grande quantidade esperada de aeronaves em voos simultâneos trará desafios para os sistemas de controle de tráfego aéreo. Além disso, estas aeronaves poderão operar sem piloto a bordo. A proposta é de que as aeronaves executem voos em rotas diretas, realizando desvios apenas quando necessário. Portanto é necessário que uma série de sistemas de detecção e resolução de conflitos atuem de forma redundante. Entre eles, está a resolução de tática de conflitos. Isso exige o desenvolvimento de ferramentas específicas para atender o novo cenário, constituído por aeronaves com características de desempenho até então inexistentes. Este trabalho investiga a possibilidade de utilização de modelos de aprendizado de reforço profundo para a solução deste problema. A detecção dos conflitos pode ser efetuada de modo independente, a partir de sistemas embarcados como o ADS-B. Após a fase de treinamento, modelos de aprendizado de reforço profundo são capazes de sugerir ações que permitam alcançar o objetivo desejado mesmo em configurações de conflitos que não tenham sido observadas previamente. Isso faz com que estes modelos sejam adequados à resolução do problema de resolução de conflitos, visto que é inviável executar um treinamento com todos os conflitos possíveis. Um sistema baseado nos modelos *Deep Q Network* foi utilizado para gerenciamento das trajetórias em caso de detecção de conflito, realizando desvios de rota com o objetivo de solucionar o conflito bem como afastar as aeronaves envolvidas o mínimo necessário de suas trajetórias ideais. Foi implementado um simulador customizado para a realização de testes utilizando diversos agentes de aprendizado de reforço profundo, bem como a comparação com estratégias alternativas. Os resultados indicam que os modelos utilizados são capazes de oferecer manobras capazes de reduzir o número de conflitos sem que as manobras realizadas para os desvios afetem de forma significativa o tempo de deslocamento ou consumo de combustível.

Palavras-chave: Mobilidade aérea urbana, Aprendizado de reforço profundo, Resolução de conflitos, Gerenciamento de Tráfego Aéreo

Abstract

Aircraft with electric propulsion and capable of performing vertical takeoff and landings, also known as eVTOL, are under development by several manufacturers and have the potential to revolutionize urban air mobility in the coming years. Adoption tends to be gradual, but once a certain level of maturity of this type of transport is reached, the expected large number of simultaneous flights will pose challenges for air traffic control systems. In addition, these aircraft are expected to be able to operate without a pilot on board. Furthermore, aircraft are supposed to fly on direct routes, making detours only when necessary. Therefore, a set of conflict detection and resolution systems is desired to act redundantly. One of those systems is the one responsible for the tactic conflict resolution. This requires developing specific tools to meet the new scenario, consisting of aircraft with performance characteristics that are not yet existent. This work investigates the possibility of using deep reinforcement learning models to solve this problem. Conflict detection can be performed independently using embedded systems as sensors, such as ADS-B. After the training phase, deep reinforcement learning models can suggest actions to achieve the desired goal even in scenarios that have not been previously observed. This capability makes these models suitable for solving the problem of conflict resolution since it is impracticable to train a system with all possible conflict configurations. A system based on Deep Q Network models was used to manage the trajectories in case of conflict detection. It carried out route deviations to resolve the conflict and deviated the aircraft involved the minimum necessary from their ideal trajectories. A customized simulator was implemented to perform tests using several deep reinforcement learning agents and compare them with alternative strategies. The obtained results indicate that the models can suggest maneuvers capable of reducing the number of conflicts without significantly affecting displacement or fuel consumption.

Keywords: Urban Air Mobility, Deep Reinforcement Learning, Conflict resolution, Air Traffic Management

Sumário

1	Introdução	1
1.1	Motivação	3
1.2	Objetivo	6
1.2.1	Objetivos Específicos	6
1.3	Estrutura da Dissertação	7
2	Conceitos Fundamentais de Inteligência Artificial	9
2.1	Processo de Decisão de Markov	9
2.2	Redes Neurais Artificiais	10
2.2.1	Retropropagação	12
2.2.2	Redes Neurais Profundas	13
2.2.3	Redes Neurais Convolucionais	14
2.3	Aprendizado de Máquina	15
2.3.1	Aprendizado de Máquina Supervisionado	16
2.3.2	Aprendizado de Máquina Não Supervisionado	17
2.3.3	Aprendizado por Reforço	18
2.3.4	Função de Recompensa	19
2.3.5	Aprendizado por Reforço Profundo	20
2.4	Modelos de Aprendizado de Reforço	21
2.4.1	<i>Q-learning</i> Básico	21
2.4.2	<i>Deep Q Learning Network</i>	22
2.4.3	<i>Double Deep Q Learning</i>	24
2.4.4	<i>Dueling Deep Q Learning</i>	24
2.4.5	<i>Proximal Policy Optimization</i>	26
2.4.6	Ator-Crítico	26
2.4.7	<i>Asynchronous Advantage Actor-Critic</i>	27
3	Conceitos Fundamentais de Gerenciamento de Tráfego Aéreo	29
3.1	OACI	29

3.2	Regras do Ar	30
3.3	Serviços de Tráfego Aéreo	32
3.3.1	Separação entre Aeronaves	35
3.3.2	Resolução de Conflitos	36
3.3.3	Voo Livre	37
3.4	Cenário Previsto para UAM	38
3.4.1	Níveis de Maturidade	38
3.4.2	Resolução de Conflitos em UAM	43
4	Revisão Bibliográfica	45
4.1	Métodos de Detecção e Resolução de Conflitos	45
4.2	Métodos de CDR Utilizando Aprendizado de Reforço Profundo	50
4.3	Discussão	54
5	Metodologia e Modelagem	56
5.1	Definição do Problema	56
5.2	Proposta de Solução	57
5.3	Metodologia	58
5.4	Modelagem	59
5.4.1	Conflitos	60
5.4.2	Eficiência dos Desvios	61
5.4.3	Função de Maximização	62
6	Implementação	63
6.1	Descrição da Implementação	63
6.2	Estratégias para Comparação	65
6.2.1	Avoid	65
6.2.2	Random	65
6.2.3	Free	65
6.3	Recompensa	66
7	Experimentos	68
7.1	Configuração	68
7.2	Ilustração de um Episódio	69
7.3	Teste de Duração do Treinamento	70
7.4	Comparação entre Agentes de DRL	72
7.4.1	Comparação com as Estratégias Alternativas	74
7.5	Densidade de Tráfego	76
7.6	Eficiência na Resolução de Conflitos	77

7.7	Discussão	78
8	Conclusão	80
8.1	Contribuições	80
8.2	Trabalhos Futuros	81
	Referências	83
	Apêndice	89
	A Produção Científica	90

Lista de Figuras

1.1	Protótipo de eVTOL	3
1.2	Espaço aéreo urbano com utilização dos eVTOL	4
1.3	Ilustração do ISUAM em ação	7
2.1	Elementos de uma rede neural artificial	11
2.2	Rede neural convolucional	15
2.3	Classificador linear	17
2.4	Agrupamento por um classificador não supervisionado.	18
2.5	Ciclo do aprendizado de reforço	19
2.6	<i>Deep Q learning Network</i>	23
2.7	Esquema de conexões entre as camadas numa rede <i>dueling</i> DQN	25
3.1	Regiões de informação de voo do Brasil.	33
3.2	Classificação dos espaços aéreos ATS.	34
3.3	Separação horizontal e vertical.	36
3.4	Conflito com perda de separação horizontal.	36
3.5	Proa relativa entre duas aeronaves em conflito.	37
3.6	Fases da resolução de conflito no UAM	43
4.1	Potencial de Voltagem Modificado (MVP)	47
4.2	Diagrama de Espaços de Solução (SSD)	48
4.3	Mudança gradual de direção e velocidade	48
4.4	Apresentação do método <i>break-circle</i>	49
4.5	Ilustração do método obstáculo velocidade (VO)	50
4.6	Arquitetura aplicada por Brittain e Wei	51
4.7	Ilustração das rotas aplicadas por Brittain e Wei	52
4.8	Tratamento de incertezas nas manobras para evitar conflitos	52
4.9	Resultados da aplicação dos modelos propostos por Li et al.	53
5.1	Fase de treinamento e aplicação do ISUAM	58

6.1	Zonas de proteção ao redor de cada aeronave.	64
6.2	Curva sugerida pela estratégia avoid para evitar um conflito.	66
7.1	Visualização de dois conflitos em um episódio.	70
7.2	Detalhes dos conflitos exibidos na Figura 7.1	71
7.3	Quantidade de episódios até a convergência do modelo.	72
7.4	Comparação entre os agentes de DRL.	73
7.5	Quantidade de intrusos na zona de conflito durante o treinamento.	75
7.6	Quantidade de intrusos na zona crítica durante o treinamento.	76
7.7	Desempenho do modelo ISUAM com variações na densidade de tráfego. . .	77
7.8	Comportamento do modelo ISUAM no cenário de baixa densidade.	78
7.9	Comparação entre as estratégias da distância percorrida no desvio.	79

Lista de Tabelas

4.1	Resumo dos Trabalhos Relacionados	54
7.1	Aeronaves presentes no episódio de ilustração	69
7.2	Comparação entre Agentes de DRL	74

Lista de Abreviaturas e Siglas

ACAS Sistema embarcado para evitar colisões, da sigla em inglês para *airborne collision avoidance system*.

ACC Centro de controle de área, da sigla em inglês para *area control center*.

APP Controle de aproximação, da sigla em inglês para *approach control*.

ATC Controle de tráfego aéreo, da sigla em inglês para *air traffic control*.

ATCo Controlador de Tráfego Aéreo, da sigla em inglês para *Air Traffic Controller*.

ATM Gerenciamento de tráfego aéreo, da sigla em inglês para *air traffic management*.

ATS Serviços de tráfego aéreo, da sigla em inglês para *air traffic services*.

CACI Convenção de Aviação Civil Internacional.

CDM Tomada de decisão colaborativa, da sigla em inglês para *Collaborative Decision Making*.

CDR Detecção e resolução de conflitos, da sigla em inglês para *conflict detection and resolution*.

CPA ponto de aproximação mais próximo, da sigla em inglês para *Closest point of approach*.

DDQN *Double Deep Q learning Network*.

DECEA Departamento de Controle do Espaço Aéreo.

DNN Rede neural profunda, da sigla em inglês para *Deep Neural Network*.

Double DQN *Double Deep Q learning Network*.

DQN *Deep Q learning Network*.

DRL Aprendizagem de reforço profunda, da sigla em inglês para *deep reinforcement learning*.

EUA Estados Unidos da América.

eVTOL Aeronaves elétricas com capacidade de decolagem e pouso na vertical.

FRA Espaço aéreo de rotas livres, da sigla em inglês para *Free Route Airspace*.

GPU Unidades de processamento gráfico, da sigla em inglês para *graphics processing unit*.

ISUAM Uma Solução de Segurança Operacional para a Mobilidade Aérea Urbana com Aprendizado de Reforço Profundo, da sigla em inglês para *Intelligent and Safe UAM with Deep Reinforcement Learning*.

MDP Processo de decisão de Markov, da sigla em inglês para *Markov decision process*.

MVP Podencial de voltagem modificado, da sigla em inglês para *Modified voltage potential*.

NASA Administração Nacional de Aeronáutica e Espaço , da sigla em inglês para *National Aeronautics and Space Administration* .

OACI Organização da Aviação Civil Internacional.

PSU Provedor de serviços UAM.

RL Aprendizado por Reforço, da sigla em inglês para *Reinforcement Learning*.

RNAV Navegação de área, da expressão em inglês para *Area navigation*.

RVSM Separação Vertical Mínima Reduzida, da sigla em inglês para *reduced vertical separation minimum*.

SSD Diagrama de espaço de soluções, da sigla em inglês para *Solution Space Diagram*.

TWR Torre de controle de aeródromo, da sigla em inglês para *aerodrome control tower*.

UAM Mobilidade aérea urbana, da sigla em inglês para *urban air mobility*.

UAV Veículo aéreo não tripulado, da sigla em inglês para *Unmanned aerial vehicle*.

UML Nível de maturidade do UAM, da sigla em inglês para *UAM Maturity Level*.

VTOL Capacidade de decolagem e pouso na vertical, da sigla em inglês para *vertical take-off and landing*.

Capítulo 1

Introdução

O transporte em grandes cidades apresenta desafios proporcionais à complexidade dos grandes centros urbanos. A mobilidade aérea urbana por meio de helicópteros já existe há várias décadas, porém tem sido utilizada apenas por uma pequena parcela da população devido aos altos custos e dificuldades na implementação de locais de embarque e desembarque. Novas aeronaves com propulsão elétrica estão em fase avançada de desenvolvimento e tem um potencial de revolucionar a mobilidade aérea urbana. Entretanto, já existe a previsão de que a facilidade de aquisição e operação destes novos veículos venha a ser motivo de preocupação em relação à segurança do espaço aéreo no futuro se alguns aspectos não forem observados [1]. O gerenciamento simultâneo de diversas aeronaves por uma entidade centralizada pode ser inviável, assim surge a necessidade de desenvolvimento de gerenciadores descentralizados de trajetórias. Esta introdução irá apresentar os desafios atuais e futuros da mobilidade aérea urbana.

Nas grandes cidades do mundo, os tempos de viagem são geralmente altos e tendem a crescer com o passar dos anos. Conseqüentemente, os destinos acessíveis em tempo aceitável estão diminuindo cada vez mais. O tempo médio de deslocamento realizado por um morador do Rio de Janeiro, por exemplo, dura 107 minutos. Uma das causas é o rápido aumento nas vendas de veículos associado ao crescimento da população desses centros urbanos sem a necessária adaptação da malha de transporte urbano [2].

No decorrer de apenas algumas décadas, as áreas urbanas em todo o mundo, tanto em países desenvolvidos como nos países em desenvolvimento, tornaram-se cada vez mais dominadas pelo automóvel e menos sustentáveis. Nos países em desenvolvimento em particular, as cidades experimentaram um rápido crescimento nos desafios relacionados ao transporte, incluindo poluição, congestionamento, acidentes, diminuição da qualidade do transporte público, degradação ambiental e mudanças climáticas, entre outros [3].

Diante deste cenário, surge a demanda para que uma parcela dos deslocamentos sejam realizados pela modalidade aérea. Assim surgiu o conceito de mobilidade aérea urbana

(UAM). Além do transporte de passageiros, a UAM abrange uma ampla variedade de conceitos operacionais que não são necessariamente a prestação do serviço de transporte como missões de emergência médica, logística ou vigilância.

A ideia de veículos voadores dentro de áreas urbanas começou na década de 1940, com o advento dos helicópteros que proporcionavam capacidade de decolagem e pouso vertical (VTOL). No início, tratava-se de uma tecnologia ainda em fase de desenvolvimento, o que levou a vários acidentes que, juntamente com altos níveis de ruído e altos custos, obrigou a maioria dos operadores nos Estados Unidos da América (EUA) e na Europa a encerrar suas operações no final da década de 1960 [4].

Até o final dos anos 1980, vários operadores começaram a fornecer os primeiros serviços de UAM usando helicópteros em Los Angeles, Nova York, São Francisco e outras cidades dos EUA. A empresa New York Airways ofereceu o serviço de passageiros entre Manhattan e o aeroporto de LaGuardia em meados da década de 1950. Esses primeiros serviços de helicóptero de passageiros eram tipicamente viabilizados por meio de uma combinação de subsídios a helicópteros e receitas do correio aéreo. Nas décadas de 1960 e 1970, a Pan Am oferecia conexões entre a região central de Manhattan em Nova York e o aeroporto internacional John F. Kennedy (JFK), permitindo que os passageiros fizessem o *check-in* no edifício da empresa 40 minutos antes da partida do voo no JFK [1].

Atualmente, os altos custos fazem com que o deslocamento urbano por meio de helicópteros em regime de fretamento exista apenas como um pequeno nicho nas grandes metrópoles como Nova York, São Paulo, Cidade do México ou Tokio [4].

O objetivo dos proponentes do UAM sob demanda em larga escala é permitir que as pessoas cheguem aos seus destinos mais rapidamente do que podem hoje em carros usando aeronaves em pelo menos um trecho da viagem. Hoje este objetivo somente pode ser atingido por meio sistema de transporte aéreo comercial quando a duração total da viagem for superior a cerca de 500 quilômetros. Para obter uma economia em viagens de até 30 quilômetros, seria necessário que os passageiros pudessem utilizar pequenas aeronaves capazes de VTOL que seriam solicitadas sob demanda, para decolar de um vertiporto no local de partida e pousar em um outro vertiporto próximo ao seu destino final. [5]

Avanços recentes na propulsão elétrica e na densidade de energia das baterias deram início a uma proliferação de projetos de aeronaves VTOL com propulsão elétrica (eVTOL). Esses veículos podem apresentar diversas formas, alguns com asas projetadas para um voo mais eficiente em cruzeiro, outros que voam com a propulsão de múltiplos rotores. Entretanto todos eles compartilham várias características importantes: decolagem e pouso verticais que permitem uma maior versatilidade operacional, propulsão elétrica ou híbrida de baixo custo e alta confiabilidade, e múltiplos rotores que são muito menores e mais



Figura 1.1: Protótipo de eVTOL denominado Volocopter 2X apresentado pela empresa Volocopter na IAA 2017 (CC BY-SA 4.0) [6]

silenciosos do que os rotores de helicóptero convencionais [7]. A Figura 1.1 apresenta o protótipo de um dos inúmeros modelos que estão sendo desenvolvidos.

Os requisitos de proficiência em pilotagem podem ser reduzidos por meio de um maior uso da automação, levando, em última análise, à possibilidade de autonomia total. Como resultado, parece haver uma promessa de que, em um futuro próximo, uma nova geração de veículos aéreos que possa realizar missões de transporte de passageiros dentro e ao redor das cidades, e de que este novo tipo de transporte será acessível o suficiente para permitir o seu uso pelo público geral [7].

A Figura 1.2 mostra a previsão de como será o espaço aéreo urbano depois que os eVTOL estiverem em pleno uso. O novo tipo de aeronave dividirá o uso dos céus com as já conhecidas das aeronaves comerciais, privadas e helicópteros.

A seguir, serão apresentados os desafios que a proliferação dos eVTOL no espaço aéreo sobre as grandes cidades irá gerar para a segurança do transporte aéreo e as propostas de solução para esse problema utilizando aprendizagem de reforço profunda.

1.1 Motivação

A possibilidade de que um grande número de aeronaves eVTOL autônomas ou pilotadas esteja disponível em um futuro próximo traz grandes preocupações em relação à segurança operacional. A motivação para o desenvolvimento deste trabalho envolve diversas



Figura 1.2: Concepção artística do espaço aéreo urbano após a plena implementação dos eVTOL

questões que surgem neste novo cenário, entre elas, a possibilidade de operação autônoma das aeronaves, a melhor eficiência da operação com rotas diretas, a possibilidade de um gerenciamento descentralizado e a utilização modelos de DRL para esse gerenciamento. A análise da literatura mostra que ainda há lacunas em relação à combinação destes temas no UAM. Portanto este trabalho buscará responder algumas das questões em aberto.

Espera-se que as primeiras operações de UAM com aeronaves eVTOL sejam realizadas para fins de demonstração e deverão cumprir as regras e regulamentos atuais do espaço aéreo. Estas regras serão muito semelhantes àsquelas aplicadas às operações atuais. Além disso, espera-se que esses voos iniciais de UAM sejam conduzidos por pilotos humanos a bordo da aeronave em todos os momentos e terão o Controle de Tráfego Aéreo (ATC) prestando serviços como nas operações atuais. No entanto, várias organizações estão buscando conceitos de UAM que abrangem a gama de funções e responsabilidades possíveis para humanos e sistemas autônomos. Por exemplo, alguns fabricantes de aeronaves UAM estão trabalhando no desenvolvimento de veículos que não exigiriam um piloto humano a bordo. Essas aeronaves seriam pilotadas remotamente por alguma combinação de humanos no solo, sistemas autônomos no solo e/ou sistemas autônomos no ar a bordo da aeronave [8]. Esta nova forma de pilotagem significa que a separação entre as rotas utilizadas pela aviação convencional também precisará ser redefinida.

As dimensões reduzidas do espaço aéreo urbano não permitem a criação de rotas predefinidas aplicando as separações convencionais entre aeronaves. Assim, será necessário

que toda a estrutura de rotas seja redefinida para proporcionar uma maior eficiência nos deslocamentos [9]. Rotas diretas proporcionam economia de tempo e reduzem as emissões de gases de efeito estufa associados aos deslocamentos. O objetivo do voo livre é permitir que cada aeronave possa voar uma rota ótima aplicando mudanças constantes na trajetória, fazendo pleno uso dos sistemas de gerenciamento embarcados. Neste tipo de voo, restrições de tráfego aéreo são impostas apenas para garantir a separação, impedir que o destino exceda a capacidade e garantir a segurança do voo. Assim, o usuário obtém máxima flexibilidade e garantia de segurança em relação à separação de outras aeronaves[10]. A eficiência proporcionada pelas rotas diretas deve estar alinhada a um gerenciamento de trajetórias compatível com o novo paradigma.

O sistema de gerenciamento de trajetórias tenderá a ser descentralizado, visto que o grande número de agentes praticamente impossibilita que todas as aeronaves sejam controladas por um único sistema. O gerenciamento das trajetórias das aeronaves evoluindo em uma determinada porção do espaço aéreo pode ser feito de forma centralizada ou descentralizada. Nos métodos centralizados, um único sistema gerencia as trajetórias para todos os agentes. No entanto, à medida que o número de aeronaves cresce, o tempo de computação desses métodos normalmente também cresce de forma exponencial. Além disso, as abordagens de planejamento centralizado precisam ser executadas novamente à medida que ocorre a modificação de qualquer informação no ambiente (por exemplo, uma nova aeronave entra no espaço aéreo, ou uma aeronave falhou em executar a trajetória que havia sido planejada). Por outro lado, os métodos descentralizados escalam melhor em relação ao número de agentes e são mais robustos uma vez que não são vulneráveis a um único ponto de falha [11]. Os requisitos para o sistema de gerenciamento descentralizado fazem com que os métodos de aprendizado de reforço sejam ideais para a solução deste problema computacional.

Um tópico central no aprendizado de máquina é a tomada de decisão sequencial. Esta é a tarefa de decidir, a partir da experiência, a sequência de ações a serem executadas em um ambiente incerto para atingir determinados objetivos [12]. Assim, o cenário descrito pode ser visto como um problema computacional, no qual existe uma quantidade infinita de configurações para os conflitos envolvendo duas ou mais aeronaves e a necessidade de definição de qual seria a melhor rota a ser tomada em cada um dos possíveis conflitos. Sistemas de aprendizagem de reforço profunda (DRL) são capazes de oferecer soluções adequadas, mesmo para situações que não tenham sido previamente observadas, desde que a distribuição do cenário real seja semelhante àquela aplicada no treinamento [13]. Esta característica torna esse tipo de modelo ideal para o desenvolvimento de um sistema de gerenciamento descentralizado de trajetórias.

Diversos trabalhos já abordaram a resolução de conflitos entre aeronaves convencionais, bem como a resolução de conflitos utilizando métodos de (DRL) também já foi estudada. Entretanto as características peculiares dos eVTOL e do espaço aéreo no qual estas aeronaves evoluirão sugere a necessidade de uma abordagem específica. Portanto, este trabalho buscará preencher esta lacuna na literatura.

1.2 Objetivo

O objetivo deste trabalho é o desenvolvimento de um sistema de gerenciamento de trajetórias de aeronaves eVTOL no ambiente UAM capaz de proporcionar voos livres garantindo a separação entre as aeronaves envolvidas. A solução proposta será denominada ISUAM - Uma Solução de Segurança Operacional para a Mobilidade Aérea Urbana com Aprendizado de Reforço Profundo, da sigla em inglês para *Intelligent and Safe UAM with Deep Reinforcement Learning*.

Este sistema utilizará DRL para oferecer uma navegação segura e eficiente, de modo que as aeronaves realizem desvios de suas trajetórias ótimas somente quando for estritamente necessário. Uma ilustração do sistema em ação é apresentada na Figura 1.3. A comparação de diferentes modelos de DRL para a resolução do problema em questão e a análise de eficiência do gerenciamento de trajetórias são os dois grandes diferenciais deste trabalho em relação aos demais já existentes na literatura.

1.2.1 Objetivos Específicos

O objetivo principal será atingido por meio dos seguintes objetivos específicos:

- Consulta ao estado da arte em termos de resolução de conflitos de tráfego aéreo utilizando DRL.
- Modelagem do problema de resolução de conflitos como um ambiente no qual o agente de DRL poderá executar ações de modo a obter recompensas quando forem realizados os desvios de forma satisfatórias.
- Criação de um ambiente onde cada modelo de DRL possa ser utilizado.
- Implementação de um simulador específico para a tarefa em questão.
- Definição dos melhores parâmetros a serem empregados em cada modelo de DRL.
- Comparação entre os resultados dos diferentes modelos.



Figura 1.3: Ilustração do ISUAM em ação. A aeronave que decola do canto direito na imagem detecta o conflito e altera sua rota para evitar o intruso que decolou do vertiporto próximo ao rio.

1.3 Estrutura da Dissertação

O restante deste texto será estruturado conforme descrito a seguir.

- No Capítulo 2 serão revistos os conceitos fundamentais de Ciência da Computação necessários para uma melhor compreensão dos métodos aplicados.
- O Capítulo 3 abordará os conceitos fundamentais de Gerenciamento do Tráfego Aéreo (ATM), considerando não somente as regras atuais como também as perspectivas do cenário futuro para a UAM.
- Os principais trabalhos relacionados serão referenciados no Capítulo 4, incluindo trabalhos que buscam a resolução de conflitos utilizando métodos diversos e também os trabalhos que aplicaram métodos de DRL para a solução de conflitos.
- No Capítulo 5 a proposta de solução para o problema será detalhada, seguida da metodologia aplicada na resolução e da modelagem do problema.
- A descrição da implementação será apresentada no Capítulo 6 acompanhada das estratégias alternativas que foram criadas para comparação com o modelo proposto.

- Os testes realizados e os resultados obtidos são expostos no Capítulo 7.
- Por fim, o Capítulo 8 traz as conclusões e as sugestões para trabalhos futuros.

Capítulo 2

Conceitos Fundamentais de Inteligência Artificial

Este capítulo descreve os principais conceitos da ciência da computação dentro do campo de inteligência artificial que serão utilizados no desenvolvimento do trabalho. Será apresentada uma introdução às redes neurais artificiais, desde a configuração mais simples até os modelos mais atuais de redes neurais convolucionais. Os conceitos de aprendizado de máquina serão revistos, incluindo a principal divisão entre aprendizado supervisionado, não supervisionado e aprendizado por reforço. Por fim, serão descritos os modelos de aprendizado por reforço baseado em *Q-learning* e também os modelos que serão aplicados de *Deep Q-learning*

2.1 Processo de Decisão de Markov

Desde a década de 1950, os Processos de decisão de Markov (MDP) têm sido estudados e aplicados em diversas disciplinas incluindo robótica, automação de controle, economia e processos de fabricação. Em um MDP, o agente pode escolher qualquer ação a que esteja disponível a partir do estado atual s_t em cada passo de tempo [14]. O processo responde na próxima etapa de tempo movendo-se para um novo estado s_{t+1} com certa probabilidade de transição e dá ao agente uma recompensa correspondente r [11].

Mais precisamente, um MDP é composto pelos seguintes componentes:

- O espaço de estados \mathcal{S} que consiste em todos os estados possíveis;
- O espaço de ações \mathcal{A} que consiste em todas as ações que o agente pode realizar;
- Função de transição $\mathcal{T}(s_{t+1}|s_t, a_t)$ que descreve a probabilidade de chegar ao estado s_{t+1} , dado o estado atual s_t e a ação a_t .

- A função de recompensa $\mathcal{R}(s_t, a_t, s_{t+1})$ que decide a recompensa imediata recebida após a transição do estado s_t para o estado s_{t+1} , devido à ação a_t . Em geral, a recompensa dependerá do estado atual, da ação atual e do próximo estado. No entanto, em alguns casos, a função de recompensa pode depender apenas do estado atual.
- Um fator de desconto $\gamma \in [0, 1]$ que decide a preferência entre a recompensa imediata versus recompensas futuras. Definir o fator de desconto menor que 1 também faz com que o sistema dê preferência para a convergência da recompensa cumulativa.

Em um problema de MDP, uma política π é um mapeamento do estado para uma ação específica (conhecida como política determinística)

$$\pi : \mathcal{S} \rightarrow \mathcal{A} \quad (2.1)$$

O objetivo do MDP é encontrar uma política ótima π^* que, caso seja seguida a partir de qualquer estado inicial, maximiza a recompensa cumulativa esperada para todos os passos futuros:

$$\pi^* = \operatorname{argmax}_{\pi} \mathbb{E} \left[\sum_{t=0}^{T-1} R(s_t, a_t, s_{t+1}) | \pi \right] \quad (2.2)$$

Função- Q e função de valor são dois conceitos importantes no MDP. A função- Q ótima $Q^*(s, a)$ é aquela cuja recompensa cumulativa esperada recebida por um agente que começa no estado s e escolhe a ação a , se somada durante todos os passos, chegará o máximo valor possível ao término da execução. Portanto, $Q^*(s, a)$ é uma indicação de quão bom é para um agente escolher a ação a enquanto está no estado s . A função de valor ótimo $V^*(s)$ denota a recompensa total máxima esperada ao iniciar do estado s , que pode ser expresso como o máximo de $Q^*(s, a)$ sobre todas as ações possíveis [11]:

$$V^*(s) = \max_a Q^*(s, a) \quad \forall s \in \mathcal{S} \quad (2.3)$$

O MDP caracterizado aqui pode ser utilizado para a formalização matemática dos modelos de aprendizado por reforço como será descrito nas seções seguintes.

2.2 Redes Neurais Artificiais

O trabalho em redes neurais artificiais (RNA), comumente chamadas apenas de “redes neurais”, foi motivado desde o início pelo reconhecimento de que o cérebro humano computa de uma maneira totalmente diferente do computador digital convencional. O cérebro

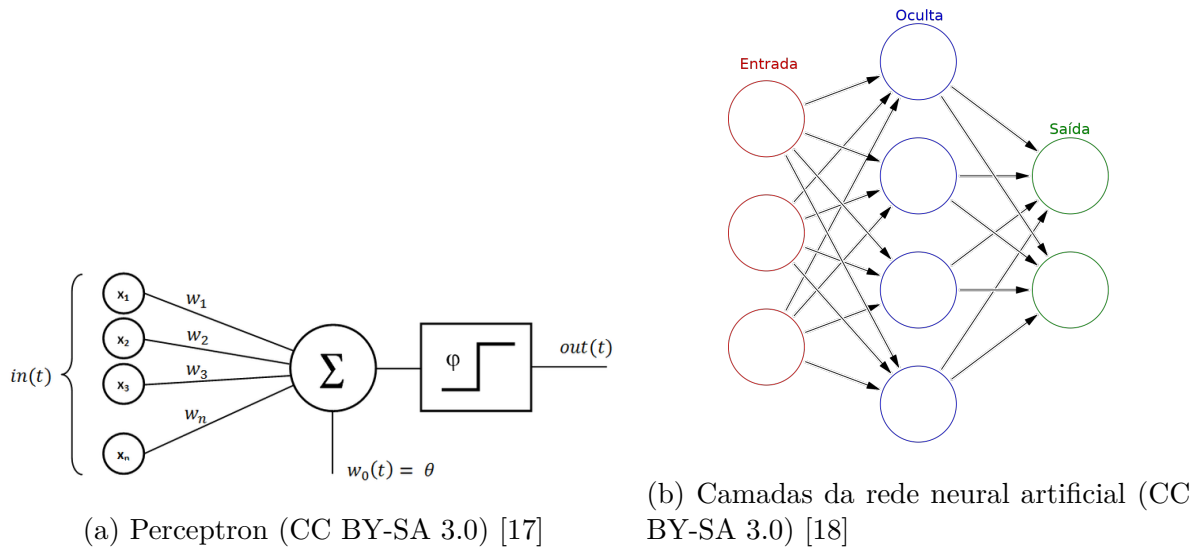


Figura 2.1: Elementos de uma rede neural artificial

é um sistema de processamento de informações altamente complexo, não linear e paralelo. Ele tem a capacidade de organizar seus constituintes estruturais, conhecidos como neurônios, de modo a realizar certos cálculos como, por exemplo, reconhecimento de padrões, percepção e controle motor, muitas vezes mais rápido do que o computador digital mais rápido existente hoje [15]. As redes neurais serão aplicadas como o principal componente do DRL desenvolvido neste trabalho.

Considerando a visão humana, que é uma tarefa de processamento de informações, é função do sistema visual fornecer uma representação do ambiente ao nosso redor e, mais importante, fornecer as informações de que precisamos para interagir com o ambiente. O cérebro rotineiramente realiza tarefas de reconhecimento perceptivo como reconhecer um rosto familiar embutido em uma cena desconhecida em aproximadamente 100-200 ms, enquanto tarefas de complexidade muito menor levam muito mais tempo em um poderoso computador [15].

As redes neurais artificiais são compostas de diversas unidades conectadas entre si denominadas *perceptrons*, conforme ilustração da Figura 2.1 (a). Cada perceptron recebe os sinais da camada imediatamente anterior x_1, x_2, \dots, x_n e calcula uma média ponderada, multiplicando cada valor pelo seu peso correspondente w_1, w_2, \dots, w_n . A esse resultado é acrescentado um valor w_0 que é constante para cada perceptron, denominado viés. Após realizar o cálculo, o perceptron poderá, ou não, propagar o resultado para as camadas seguintes, a depender do resultado da aplicação da chamada função de ativação ϕ sobre o resultado obtido anteriormente [16].

Os perceptrons são organizados em camadas, como mostra a Figura 2.1 (b). A camada de entrada recebe os sinais que serão processados. Nas redes neurais simples poderá existir

ou não uma camada oculta. A quantidade de perceptrons nas camadas de entrada e saída depende de quantos sinais deseja-se receber ou produzir, respectivamente.

Se o problema de classificação for separável, ainda é necessário definir os pesos para cada perceptron, de modo que a rede resolva corretamente o problema de classificação. Isso pode ser feito de maneira iterativa, apresentando exemplos com classificações conhecidas, um após o outro. Esse processo é chamado de aprendizado ou treinamento, pois se assemelha ao processo pelo qual passamos ao aprender alguma coisa. A simulação da aprendizagem por computador envolve pequenas alterações nos pesos e no limiar cada vez que um novo exemplo é apresentado de forma que a classificação seja melhorada. [19].

2.2.1 Retropropagação

O algoritmo retropropagação aprende os pesos para uma rede multicamadas, considerando uma rede com um conjunto fixo de unidades e interconexões. Ele emprega gradiente descendente para tentar minimizar o erro quadrado entre os valores de saída da rede e os valores de destino para essas saídas. Esta subseção apresenta o algoritmo retropropagação, o modelo mais simples de aprendizado de uma rede neural, em seguida serão abordados modelos e métodos mais sofisticados.

Considerando uma rede com várias unidades de saída, E é definido como a soma dos erros de todas as unidades de saída da rede [20]:

$$E(\vec{w}) = 1/2 \sum_{d \in D} \sum_{k \in \text{saídas}} R(t_{kd} - o_{kd})^2 \quad (2.4)$$

Onde *saídas* é o conjunto de unidades de saídas na rede, e t_{kd} e o_{kd} são, respectivamente, os valores alvo e de saída associados com a k -ésima unidade de saída e ao exemplo de treinamento d .

O problema de aprendizado enfrentado pelo algoritmo de retropropagação é buscar um grande espaço de hipóteses definido por todos os valores de peso possíveis para todas as unidades da rede. O erro pode ser interpretado como uma superfície multidimensional onde as outras dimensões do espaço correspondem agora a todos os pesos associados a todas as unidades da rede. Como no caso de treinar uma única unidade, o gradiente descendente pode ser usado para tentar encontrar uma hipótese para minimizar E , ou seja, encontrar um mínimo local nessa superfície. Uma grande diferença no caso de redes multicamadas é que a superfície de erro pode ter múltiplos mínimos locais. Infelizmente, isso significa que um gradiente descendente garante apenas que o sistema convergirá para algum mínimo local, e não necessariamente para o erro de mínimo global. Apesar desse obstáculo, na prática, o algoritmo de retropropagação tem produzido excelentes resultados em muitas aplicações do mundo real [20].

O algoritmo descrito aqui se aplica a redes *feedforward* em camadas contendo duas camadas de unidades sigmóides, com unidades em cada camada conectadas a todas as unidades da camada anterior. Esta é a versão de gradiente descendente incremental ou estocástica de retropropagação.

O algoritmo de retropropagação começa construindo uma rede com o número desejado de unidades ocultas e de saída e inicializando todos os pesos da rede com pequenos valores aleatórios. Dada essa estrutura de rede fixa, o laço principal do algoritmo itera repetidamente sobre os exemplos de treinamento. Para cada exemplo de treinamento, ele aplica a rede ao exemplo, calcula o erro da saída da rede para este exemplo, calcula o gradiente em relação ao erro neste exemplo, então atualiza todos os pesos na rede. Essa etapa de descida do gradiente é iterada (geralmente milhares de vezes, usando os mesmos exemplos de treinamento várias vezes) até que a rede tenha um desempenho aceitável [20].

2.2.2 Redes Neurais Profundas

Em visão computacional no último meio século, pesquisas significativas foram feitas para encontrar os melhores parâmetros em sistemas para classificação precisa, e muitos filtros de imagem, transformações e convoluções diferentes foram propostos para implementar esses extratores de recursos manualmente.

Embora essas abordagens tenham tido algum sucesso, algoritmos de aprendizado estão alcançando maior precisão recentemente com *big data* e computadores poderosos. Com poucas suposições e pouca interferência manual, estruturas semelhantes ao cone hierárquico estão sendo aprendidas automaticamente a partir de grandes quantidades de dados. Essas abordagens de aprendizado são especialmente interessantes porque aprendem, não são voltadas para nenhuma tarefa específica e podem ser usadas em uma variedade de aplicações. Os novos modelos são capazes de aprender os extratores de recursos ocultos e também como melhor combiná-los para definir a saída [12].

Essa é a ideia por trás das redes neurais profundas, onde, a partir da entrada bruta, cada camada oculta combina os valores da camada anterior e aprende funções mais complexas a partir da entrada. O fato de que os valores das unidades ocultas são contínuos permite uma representação mais refinada e graduada de entradas semelhantes.

Uma rede neural profunda é caracterizada por uma sucessão de múltiplas camadas de processamento. Cada camada consiste em uma transformação não linear e a sequência dessas transformações leva ao aprendizado de diferentes níveis de abstração.

2.2.3 Redes Neurais Convolucionais

Redes neurais convolucionais (CNNs) são análogas às redes neurais artificiais tradicionais, pois são compostas por *perceptrons*, cujo objetivo é simular o comportamento dos neurônios biológicos, que se auto-otimizam por meio do aprendizado. Cada *perceptron* ainda receberá uma entrada e realizará uma operação como, por exemplo, um produto escalar seguido por uma função não linear. Desde os vetores de imagem bruta de entrada até a saída final da pontuação de cada classe, toda a rede ainda expressará uma única função de pontuação perceptiva. Trata-se de um tipo mais sofisticado de rede neural que é aplicado por muitos dos modelos de DRL utilizados neste trabalho.

A última camada conterá funções de perda associadas às classes, e todas soluções que já foram desenvolvidas para RNAs tradicionais ainda se aplicam. Uma diferença notável entre as CNNs e as RNAs tradicionais é que as RNAs são usadas principalmente no campo de reconhecimento de padrões em imagens. Isso nos permite codificar recursos específicos da imagem na arquitetura, tornando a rede mais adequada para tarefas focadas em imagens, enquanto reduz ainda mais os parâmetros necessários para configurar o modelo [21].

Uma das maiores limitações das formas tradicionais de RNA é que elas tendem a lutar com a complexidade computacional necessária para computar dados de imagem. Conjuntos de dados comuns de análise comparativa para o aprendizado de máquina, como o banco de dados MNIST de dígitos manuscritos [22], são adequados para a maioria das formas de RNA, devido à sua dimensionalidade de imagem relativamente pequena de apenas 28×28 . Com esse conjunto de dados, um único neurônio na primeira camada oculta conterá 784 pesos ($28 \times 28 \times 1$), considerando que cada entrada do MNIST é normalizada apenas para valores preto e branco, o que é gerenciável para a maioria das formas de RNA. Se for considerada uma entrada de imagem colorida mais substancial de 64×64 , o número de pesos em apenas um único neurônio da primeira camada aumenta substancialmente para 12.288. Levando em consideração também que para lidar com essa escala de entrada, a rede também precisa ser muito maior do que a rede usada para classificar dígitos MNIST com cores normalizadas, então será possível compreender as desvantagens de usar tais modelos [21].

Conforme observado anteriormente, as CNNs são construídas sob a premissa de que a entrada será composta de imagens. Isso foca na arquitetura a ser configurada de forma a melhor atender a necessidade de lidar com esse tipo específico de dados. Uma das principais diferenças é que os neurônios que compõem as camadas dentro da CNN são organizados em três dimensões, a dimensionalidade espacial da entrada (altura e largura) e a profundidade. A profundidade não se refere ao número total de camadas dentro da RNA, mas sim à terceira dimensão de um volume de ativação. Ao contrário do RNAs

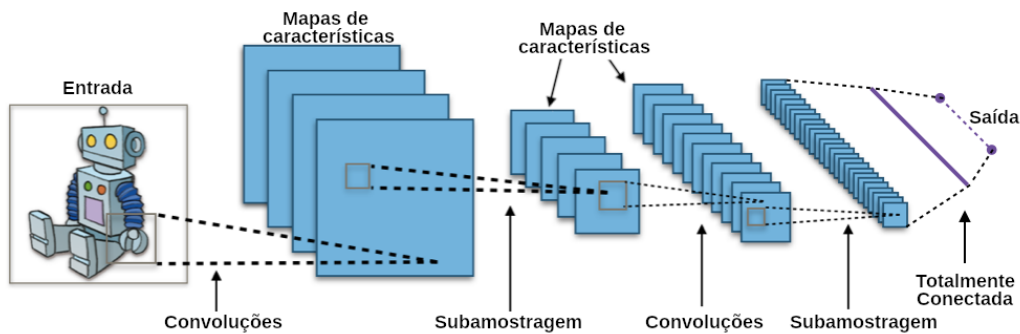


Figura 2.2: Típica rede neural convolucional (CC BY-SA 4.0) [23]

padrão, os neurônios dentro de qualquer camada se conectarão apenas a uma pequena região da camada anterior [21].

Na prática, isso significaria que, para o exemplo dado anteriormente, os dados de entrada terão uma dimensionalidade de $64 \times 64 \times 3$ (altura, largura e profundidade), levando a uma camada de saída final composta por uma dimensionalidade de $1 \times 1 \times n$, onde n representa o número possível de classes, pois teríamos condensado a dimensionalidade de entrada completa em um volume menor de classes [21].

2.3 Aprendizado de Máquina

O aprendizado de máquina pode ser amplamente definido como um conjunto de métodos computacionais que utilizam a experiência passada, ou seja, o aprendizado, para melhorar o desempenho futuro ou fazer previsões mais precisas quando deparados com novas instâncias do mesmo tipo de problema. Neste caso, o termo experiência refere-se à informação passada disponível para o sistema que está aprendendo, que normalmente toma a forma de dados coletados e disponibilizados para análise. Esses dados podem estar sob a forma de conjuntos de treinamento rotulados por humanos ou outros tipos de informações obtidas através da interação com o meio ambiente. Em todos os casos, sua qualidade e tamanho são cruciais para o sucesso das previsões feitas pelo sistema que está aprendendo [24]. O aprendizado por reforço será descrito a seguir como uma das categorias do aprendizado de máquina.

O aprendizado de máquina consiste em projetar algoritmos que sejam capazes de realizar previsões eficientes e precisas. Como em outras áreas da ciência da computação, algumas medidas críticas da qualidade desses algoritmos são sua complexidade de tempo e espaço de armazenamento. Porém, no aprendizado de máquina é necessário adicionalmente que seja conhecido o tamanho da amostra necessária para que o algoritmo aprenda uma família de conceitos para que se possa fazer uma avaliação mais precisa da complexidade do sistema. De maneira mais geral, a qualidade do aprendizado de um algoritmo

depende da complexidade das classes consideradas e o tamanho da amostra de treinamento [24].

Como o sucesso de um algoritmo de aprendizado depende dos dados utilizados, o aprendizado de máquina está inerentemente relacionado à análise de dados e estatísticas. Em outras palavras, as técnicas de aprendizado são métodos orientados a dados que combinam conceitos fundamentais em ciência da computação com ideias de estatística, probabilidade e otimização [24].

As técnicas de aprendizado de máquina são basicamente divididas em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. No aprendizado supervisionado, o agente observa exemplos, que são apresentados na forma de pares de entrada e saída e aprende uma função que será capaz de mapear as respostas dadas às entradas que são apresentadas. A parte fundamental é a rotulagem dos dados de treinamento que servirão para formar o aprendizado da máquina. No aprendizado não supervisionado, o agente aprende padrões analisando as entradas apresentadas, sem receber qualquer tipo de orientação em relação à qualidade da classificação. Por fim, no aprendizado por reforço, um agente interage com o ambiente com o objetivo de determinar quais ações são necessárias para maximizar a recompensa recebida [25]. Abaixo cada uma das categorias será examinada em mais detalhes.

2.3.1 Aprendizado de Máquina Supervisionado

Quando existe a disponibilidade de um grande conjunto de dados previamente rotulados, o aprendizado de máquina supervisionado poderá ser aplicado. Os resultados tendem a ser melhores quando os dados a serem classificados possuem uma segmentação bem definida. Esse sistema de aprendizado requer uma prévia rotulagem dos dados, ou seja, que cada elemento que faz parte do banco de dados de treinamento contenha uma anotação indicando a qual classe aquele elemento pertence. Em seguida é realizado o treinamento do sistema a partir do conjunto de dados marcados. Uma vez que o modelo esteja treinado, ele será capaz de reconhecer novos elementos e classificá-los em uma das categorias pré-definidas com base nas semelhanças entre o novo elemento e os elementos de cada uma das classes utilizadas no treinamento [26].

Como exemplo de aprendizado supervisionado, podemos citar um sistema que busca realizar uma classificação linear, ou seja, busca os parâmetros para uma reta no plano cartesiano, que seja capaz oferecer a melhor classificação entre dois grupos de pontos. À medida que o treinamento é realizado, os parâmetros da reta são atualizados de modo que o menor número possível de pontos seja classificado fora do seu grupo esperado. Caso o objetivo seja segmentar os pontos presentes na Figura 2.3, após a observação de algumas

dezenas de pontos o classificador é capaz de definir uma linha vermelha que divide os dois grupos.

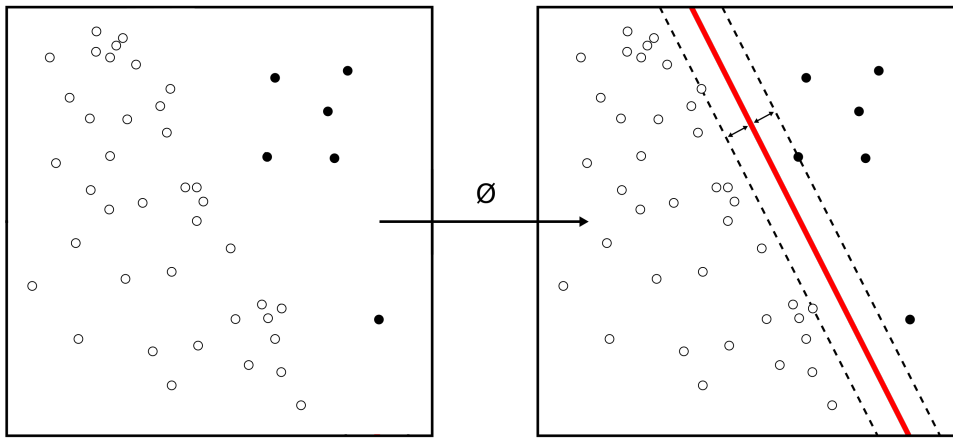


Figura 2.3: Grupos de dados que podem ser facilmente separados por um classificador linear

2.3.2 Aprendizado de Máquina Não Supervisionado

O aprendizado não supervisionado tenta extrair informações de dados sem rótulos, por exemplo, agrupamento e estimativa de densidade. O aprendizado por representação é um tipo clássico de aprendizado não supervisionado. Ela busca uma representação que seja capaz de preservar o máximo de informações possível sobre os dados originais, ao mesmo tempo, para manter a representação mais simples ou mais acessíveis do que os dados originais, com representações de baixa dimensão, esparsas e independentes [27].

Os métodos de aprendizado de máquina não supervisionados são particularmente úteis em tarefas de descrição porque visam encontrar relacionamentos em uma estrutura de dados sem ter um resultado medido. Essa categoria de aprendizado de máquina é chamada de não supervisionada porque não possui uma variável de resposta que possa supervisionar a análise. O objetivo do aprendizado não supervisionado é identificar dimensões, componentes, agrupamentos ou trajetórias subjacentes dentro de uma estrutura de dados [28].

Existem diversos domínios onde há uma grande quantidade de dados disponíveis, porém estas informações não estão classificadas. A análise de trajetórias de chegadas e saídas em aeroportos, por exemplo, pode contribuir significativamente para melhorar as operações futuras por meio de duas formas principais: análise posteriores do desempenho operacional para identificar ineficiências e fornecer orientação nos ajustes do sistema e caracterização do comportamento real do sistema que alimenta os modelos orientados para suporte à decisão em tempo real. Nesse caso, um algoritmo de agrupamento baseado em

densidade pode ser utilizado usado para identificar os principais padrões de trajetória de voo no espaço aéreo [29].

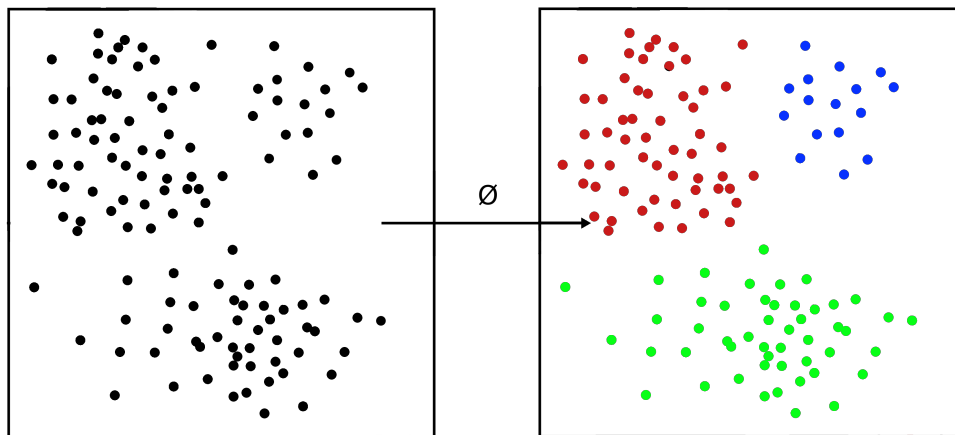


Figura 2.4: Agrupamento por um classificador não supervisionado.

O exemplo da Figura 2.4 mostra um conjunto de dados recebido pelo sistema. A figura do lado direito mostra um possível agrupamento em torno de três núcleos distintos. Existem diversas formas para classificar o erro de um determinado agrupamento e cada uma delas pode gerar resultados diferentes.

2.3.3 Aprendizado por Reforço

O objetivo da aprendizagem por reforço é definir o que fazer em cada situação, em outras palavras, como mapear situações para ações, de modo a maximizar um sinal numérico de recompensa. O sistema que está aprendendo não é orientado sobre quais ações tomar, como na maioria das formas de aprendizado de máquina, mas, em vez disso, deve descobrir qual as ações rendem a maior recompensa ao serem tomadas. No mais interessante e desafiador dos casos, as ações podem afetar não apenas a recompensa imediata, mas também a próxima situação e, através disso, todas as recompensas subsequentes. Essas duas características: 1) tentativa e erro; e 2) busca e recompensa atrasada são as duas características distintivas mais importantes de aprendizado por reforço [30].

A aprendizagem por reforço é definida não pela caracterização dos métodos de aprendizagem, mas pela caracterização do problema de aprendizagem. Qualquer método que seja adequado para resolver esse problema, consideramos ser um método de aprendizado por reforço. A ideia básica é simplesmente capturar os aspectos mais importantes do problema real enfrentado por um agente de aprendizagem interagindo com seu ambiente para atingir um objetivo. Claramente, tal agente deve ser capaz de perceber o estado do meio ambiente até certo ponto e deve ser capaz de tomar ações que afetam o estado.

O agente também deve ter um objetivo ou objetivos relacionados ao estado do ambiente [30].

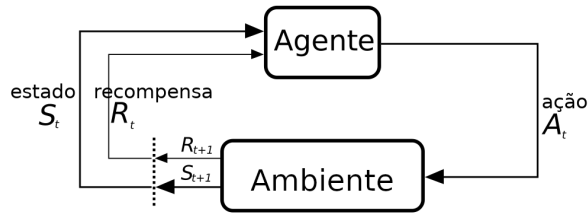


Figura 2.5: Ciclo do aprendizado de reforço (CC BY-SA 4.0) [31]

Deseja-se, no aprendizado por reforço, que o agente seja capaz de prever quão boa qualquer ação intermediária é para levá-lo à recompensa real. Essa é estimativa interna de recompensa para a ação. Inicialmente, essa estimativa de recompensa para todas as ações é zero porque o sistema ainda não possui informação. São necessários dados para aprender, então o sistema começa a fazer a exploração na qual poderá tomar certas ações e observar a recompensa recebida em troca. Em seguida, o sistema atualiza suas estimativas internas usando essas informações [?].

À medida que o agente continua explorando, ele coleta mais dados e aprende mais sobre o ambiente e quão boas são suas ações. Quando é alcançado um nível em que as estimativas de recompensa de ações são boas o suficiente, inicia a fase em que o sistema passa a tirar proveito do que já foi aprendido. Isso é feito tomando as ações que geram a maior recompensa de acordo com a estimativa interna de recompensa. No início, quando o agente não sabe muito sobre o ambiente, ele experimenta ações ao acaso. À medida que ele aprende mais, passa gradualmente de exploração para aproveitamento, passando de escolhas aleatórias para aquelas influenciadas pelas estimativas internas de recompensa [?].

2.3.4 Função de Recompensa

A escolha de uma abordagem de aprendizado por reforço permite que teóricos e profissionais se concentrem na maximização eficiente, flexível e eficaz de sinais de recompensa configurados arbitrariamente em ambientes arbitrários, deixando de lado o design e a implementação de sinais de recompensa para mais tarde. O formalismo de aprendizado por reforço é um enquadramento útil do problema de IA na medida em que simplifica a tarefa dos pesquisadores, fornece a eles uma estrutura comum para comunicar técnicas e resultados e concentra o trabalho nas partes frutíferas e interessantes do problema enquanto afasta a atenção de partes menos importantes. À medida que os sistemas de IA baseados

em aprendizado por reforço se tornam mais gerais e autônomos, o design de mecanismos de recompensa que provocam comportamentos desejados torna-se mais importante e mais difícil [32].

A exploração é amplamente considerada como o desafio mais significativo em aprendizado de reforço, porque o agente pode ter que executar sequências precisas de ações para alcançar estados com alta recompensa. Aqui, estratégias simples de exploração aleatória provavelmente falham: por exemplo, uma caminhada aleatória pode levar um tempo exponencial para alcançar o canto do ambiente onde o agente pode acumular alta recompensa. Embora o aprendizado por reforço tenha visto um tremendo aumento da atividade de pesquisa recente, essencialmente todos os algoritmos padrão implantados na prática empregam randomização simples ou suas variantes e, conseqüentemente, incorrem em complexidade de amostra extremamente alta [33].

A definição de uma boa função de recompensa é muito importante para um sistema inteligente artificial voltado para ATM. Diferentes estruturas de recompensa resultarão em diferentes desempenhos do sistema. Trata-se de um fator que tem impacto direto no processo de aprendizagem. Uma função de recompensa bem definida avalia um cenário e aconselha a melhor ação em uma situação apropriada. Assim, é necessário criar funções de recompensa especiais para alguns assuntos e/ou situações reais específicas para alcançar melhores resultados. A função de recompensa precisa ser treinada no ambiente que irá atuar, assim seus resultados podem ser aprimorados por meio desse processo [34].

2.3.5 Aprendizado por Reforço Profundo

A utilização de redes neurais profundas permite que o aprendizado de reforço seja dimensionado para problemas de tomada de decisão que antes eram intratáveis, ou seja, configurações com estado de alta dimensão e espaços de ação. Entre os trabalhos recentes na área de aprendizado de reforço profundo (DRL), há duas histórias de sucesso notáveis. O primeiro, iniciando a revolução no DRL, foi o desenvolvimento de um algoritmo que pudesse aprender a jogar uma variedade de títulos de jogos de videogame feitos para o console Atari 2600 em um nível sobre-humano, apenas tendo como entrada os pixels da imagem. Fornecendo soluções para a instabilidade das técnicas de aproximação de função em aprendizado de reforço, este trabalho foi o primeiro a demonstrar de forma convincente que os agentes de aprendizado de reforço podem ser treinados em observações brutas de alta dimensão, exclusivamente com base em um sinal de recompensa. O segundo sucesso de destaque foi o desenvolvimento de um sistema DRL híbrido, AlphaGo, que derrotou um campeão mundial humano em Go, paralelamente à conquista histórica do Deep Blue da IBM no xadrez duas décadas antes e do sistema Watson DeepQA da IBM que venceu aqueles que eram considerados os melhores jogadores humanos no pro-

grama Jeopardy!, um programa de TV de perguntas e respostas. Ao contrário das regras artesanais que dominaram os sistemas de jogo de xadrez, o AlphaGo era composto por redes neurais treinadas usando aprendizado supervisionado e por reforço, em combinação com um algoritmo de busca heurística tradicional [35].

O algoritmo de aprendizado por reforço mais conhecido que usa redes neurais (mas não redes profundas, ou seja, há apenas uma camada oculta) é o jogador de gamão chamado TD-Gammon, que ganhou uma pontuação igual à dos campeões humanos ao jogar contra si mesmo. TD-Gammon usa o algoritmo TD lambda para treinar uma rede neural simples para aprender a jogar gamão. No entanto, tentativas posteriores de usar o mesmo método para outros jogos, como xadrez, Go e damas, não obtiveram sucesso. Com o crescente interesse em trabalhos de pesquisa em aprendizado profundo em meados da década de 2000, surgiu a promessa de usar redes neurais como aproximador de função tanto para a função de valor de estado $V(s)$ quanto para a função de valor de ação $Q(s, a)$ em tarefas visuais de aprendizado de reforço. Nas subseções seguintes, serão apresentados os trabalhos que usaram as redes neurais profundas em combinação com a estrutura de aprendizado por reforço para melhorar o desempenho das políticas de controle de aprendizado, enfatizando aquelas que são alimentadas com dados de entrada brutos [36].

2.4 Modelos de Aprendizado de Reforço

Esta seção descreve os modelos de aprendizado de reforço. Inicialmente será abordado *Q-learning* básico, um dos modelos mais simples. Na sequência serão apresentadas as suas evoluções que serão os modelos utilizados neste trabalho.

2.4.1 *Q-learning* Básico

Em contraste com os algoritmos anteriores que não diferenciavam o comportamento da aprendizagem, o *Q-learning* usa um método para separar a política de atuação da política de aprendizagem [34]. Como resultado, mesmo que a ação selecionada no próximo estado seja medíocre, a informação não foi incluída na atualização da função Q do estado atual, e o dilema é que se trata de uma escolha errada. No entanto, com a separação das políticas o *Q-learning* resolve o dilema. A equação para o valor Q é a seguinte [20]:

$$Q(s, a) \leftarrow Q(s, a) + \alpha [R + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (2.5)$$

onde α é a taxa de aprendizado e tem um valor entre 0 e 1. R é uma recompensa e γ é a taxa de redução da recompensa com o passar do tempo.

O valor-Q ($Q(S, A)$) da ação para o estado atual S é atualizado com a soma do valor existente $Q(S, A)$ e a equação que determina a melhor ação no estado atual. *Q-learning* continua atualizando o valor-Q para cada estado continuamente usando a equação acima. Antes de iniciar o *Q-learning*, as recompensas estão presentes na tabela Q . Se um agente selecionar uma ação por meio de uma política no estado inicial, ele passará para o próximo estado usando a função de transição aplicável ao modelo. Este processo é repetido várias vezes até que o valor-Q geral convirja para um valor específico onde a tabela Q é usada para resolver um determinado problema. *Q-learning* combina programação dinâmica e métodos de Monte Carlo, que têm sido usados para resolver a equação de Bellman [37]:

$$Q^\pi(s, a) = \sum_{s' \in S} T(s, a, s')(R(s, a, s') + \gamma Q^\pi(s', a = \pi(s')))) \quad (2.6)$$

Essa abordagem se tornou a base de muitos algoritmos de aprendizado por reforço porque, ao contrário de outros métodos, o *Q-learning* é simples e exibe uma excelente capacidade de aprendizado em ambientes onde há um único agente. No entanto, no *Q-learning*, um valor é atualizado apenas uma vez por ação. Portanto, é difícil resolver efetivamente problemas complicados em um ambiente onde há um grande número de pares de estado-ação porque muitos pares podem não ter sido experimentados anteriormente. Além disso, como a tabela Q para recompensas é predefinida, é necessária uma quantidade consideravelmente grande de memória de armazenamento. Em um ambiente multi-agente com dois ou mais agentes, é necessária uma grande memória de estado-ação, o que leva a problemas.

2.4.2 *Deep Q Learning Network*

Para solucionar alguns dos principais problemas do *Q-learning* Básico, foi desenvolvida uma abordagem chamada *Deep Q learning Network* (DQN) [38], que se beneficia das vantagens do aprendizado profundo para representação abstrata na política ótima de aprendizado. Ou seja, trata-se de um modelo capaz de selecionar ações de forma a maximizar o valor esperado da soma cumulativa de recompensas. O DQN combina uma rede neural convolucional profunda com o método de aprendizado por reforço mais simples (*Q-learning*) para jogar vários jogos de computador Atari 2600 apenas observando a tela.

Porém, a combinação de algoritmos de aprendizado por reforço sem modelo como um mero algoritmo de *Q-learning* com redes neurais causa alguns problemas de estabilidade que fazem com que os resultados sejam divergentes. Existem duas razões principais para esses problemas, a primeira delas é o fato de que os estados subsequentes nas tarefas de aprendizado de reforço estão muito correlacionados. A segunda razão é que a política está mudando com frequência, devido a pequenas alterações nos valores Q .

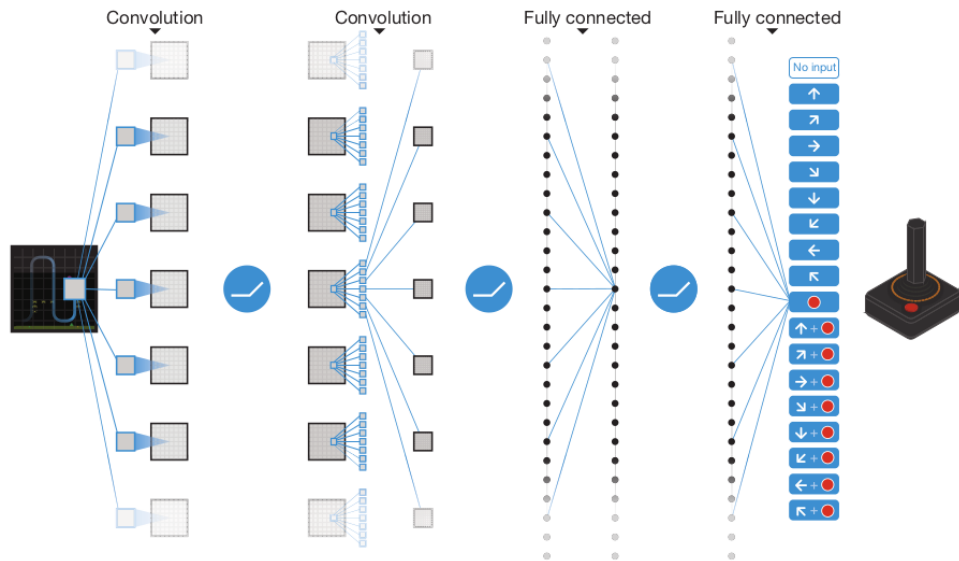


Figura 2.6: *Deep Q learning Network* .

Para lidar com esses problemas, a DQN fornece as soluções a seguir. Para a questão dos estados correlacionados, ele utiliza uma abordagem chamada de experiência de rejogo. No processo de aprendizado, DQN armazena a experiência do agente (s_t, a_t, r_t, r_{t+1}) a cada passo de tempo em um conjunto de dados D , onde s_t , a_t e r_t , respectivamente, o estado, a ação selecionada e a recompensa recebida no intervalo de tempo t e s_{t+1} é o estado no próximo intervalo de tempo. Para atualizar os valores Q , ele usa atualizações estocásticas de mini lote com amostragem aleatória uniforme da memória de rejogo da experiência (transições anteriores) no tempo de treinamento. Este trabalho quebra fortes correlações entre amostras consecutivas, e para instabilidade na política, a rede é treinada com uma rede Q alvo para obter alvos Q-learning consistentes fixando os parâmetros de peso usados no alvo Q-learning e atualizando-os periodicamente [36].

Até recentemente o método proposto alcançava figurava entre os melhores agentes de DRL. Em alguns jogos, sua estratégia superou o desempenho do jogador humano e alcançou resultados de ponta em muitos jogos Atari com a mesma arquitetura de rede ou hiperparâmetros. Vários fatores são responsáveis pelos resultados significativos obtidos. Primeiro, os avanços no poder de computação, especialmente a tecnologia de unidades de processamento gráfico (GPU) altamente paralelas, que permitiu o treinamento de redes neurais profundas com milhares de parâmetros de peso. Em segundo lugar, o DQN usou uma CNN grande e profunda, que tornou o aprendizado de representação melhor. Em terceiro lugar, o DQN usou replay de experiência para o problema de estados correlacionados [36].

2.4.3 *Double Deep Q Learning*

O Double Q-learning foi desenvolvido para resolver um problema apresentado pelo Q-learning de não funcionar bem em um ambiente estocástico. Neste ambiente, o Q-learning é tendencioso porque o valor da ação do agente é superestimado. O Q-learning convencional não procura nenhum novo valor após um certo tempo, mas seleciona repetidamente o valor mais alto entre os valores existentes. Assim, Hasselt *et al.* [39] desenvolveu o *Double Q learning*, que resolve oferece uma abordagem melhor para o problema de Q-learning ao dividir a função de avaliação que determina a ação para evitar o desvio do valor no algoritmo do Q-learning. O algoritmo existente é o mesmo do Q-learning. A Equação 2.5 é dividida em duas equações, e o valor é derivado de forma seletiva e aleatória [37]:

$$Q^A(s, a) \leftarrow Q^A(s, a) + \alpha(s, a)(R + \gamma Q^B(s', a') - Q^A(s, a)) \quad (2.7)$$

$$Q^B(s, a) \leftarrow Q^B(s, a) + \alpha(s, a)(R + \gamma Q^A(s', a') - Q^B(s, a)) \quad (2.8)$$

Q-learning duplo tem duas funções Q e cada uma delas é atualizada com o valor da outra função Q. As duas funções Q são importantes para aprender a partir de um conjunto separado de experiências, mas ambas as funções de valor são usadas para escolher a ação. O *Double Q-learning* foi desenvolvido ativamente e combinado com o deep Q-learning para desenvolver o double deep Q-learning (Double DQN). O Q-learning duplo profundo também melhorou o desempenho do Q-learning profundo, evitando previsões otimistas e divergências de valores Q que expressam valores futuros.

2.4.4 *Dueling Deep Q Learning*

Os modelos de aprendizado profundo abordados até este ponto consideram arquiteturas sequenciais. Nesses modelos, todos os neurônios em qualquer camada específica podem ser conectados apenas aos neurônios em apenas uma camada antes e uma camada após sua própria camada. Em outras palavras, não existiam divisões ou laços nas arquiteturas dos modelos.

Embora o DQN e o *Double DQN* tenham duas redes Q, nessas arquiteturas há apenas um modelo de aprendizado profundo e os outros valores da rede alvo são uma cópia periódica dos valores da rede ativa. No *dueling DQN*, há uma arquitetura não sequencial de aprendizado profundo, no qual, após as camadas convolucionais, as camadas do modelo ramificam-se em dois fluxos diferentes (sub-redes), cada um com sua própria camada totalmente conectada e camadas de saída. O primeiro desses dois ramos corresponde ao da função de valor que é usado para estimar o valor de um determinado estado e possui

um único nó em sua camada de saída. O segundo ramo é chamado de rede de vantagem e calcula o valor da vantagem de tomar uma ação específica sobre o valor base de estar na corrente estado [40].

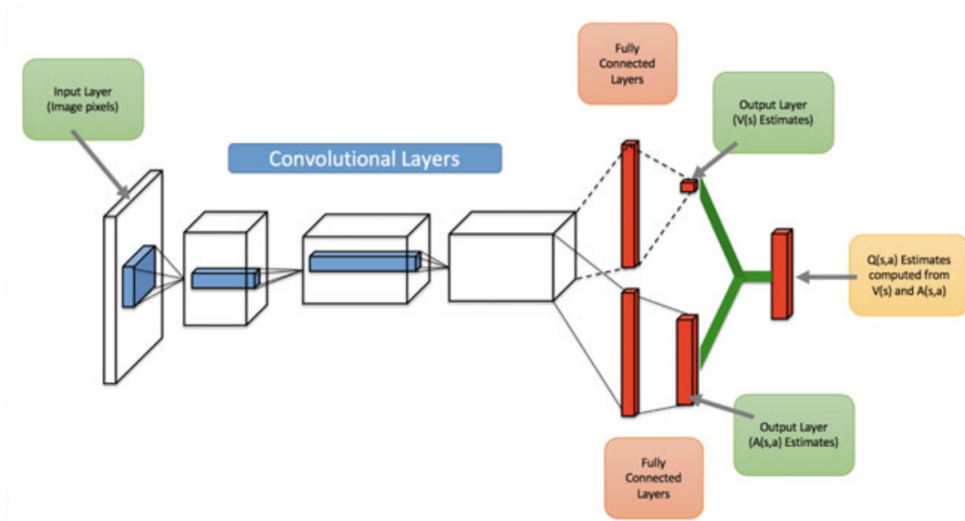


Figura 2.7: Esquema de conexões entre as camadas numa rede *dueling* DQN .

A função Q no *dueling* DQN ainda representa a função Q em qualquer outro algoritmo de aprendizado Q e, portanto, o algoritmo *dueling* DQN [41] deve funcionar da mesma maneira conceitualmente que o algoritmo de aprendizado Q estimando os valores de ação absolutos ou as estimativas Q . Então é necessário estimar os valores para os pares de ação-valor/ Q também. O par ação-valor é o valor absoluto de tomar uma determinada ação em um determinado estado. Portanto, se for possível combinar a saída do valor base do estado com os valores incrementais de vantagem das ações da segunda rede de vantagem, então seria possível estimar o valor para o par ação-valor conforme necessário no aprendizado Q . Isso pode ser representado matematicamente como [40]:

$$Q_{(s,a;\theta,\alpha,\beta)} = V_{(s;\theta,\beta)} + (A_{(s,a;\theta,\alpha)} - \max_{a' \in |A|} A_{(s,a';\theta,\alpha)}) \quad (2.9)$$

Na Equação acima dos termos Q , V , s , a e a' têm o mesmo significado discutido anteriormente neste texto. Além disso, o termo A indica o valor de vantagem, θ representa o vetor de parâmetro da camada convolucional, que é comum às redes valor vantagem. α representa o vetor de parâmetro da rede vantagem e β representa o vetor de parâmetro da função estado-valor. Como entramos no domínio dos aproximadores da função, os valores de qualquer rede são indicados em relação aos parâmetros da rede de “estimativa” para distinguir entre os valores/estimativas da mesma variável estimada a partir de múltiplas estimativas diferentes funções.

2.4.5 *Proximal Policy Optimization*

Schulman *et al.* [42] propôs uma família de métodos de gradiente de política para aprendizado por reforço denominada *Proximal Policy Optimization* (PPO), que alterna entre a amostragem de dados por meio da interação com o ambiente e a otimização de uma função objetivo “substituta” usando gradiente estocástico ascendente. Considerando que os métodos de gradiente de política padrão executam uma atualização de gradiente por amostra de dados, este método propôs uma nova função objetivo que permite várias épocas de atualizações de subconjunto de exemplos, também chamados de *minibatch*.

A diferença entre métodos baseados em valor e baseados em políticas é essencialmente uma questão de onde reside o ônus da otimização. Ambos os tipos de métodos devem propor ações e avaliar o comportamento resultante, mas enquanto os métodos baseados em valor se concentram na avaliação da recompensa cumulativa ideal e têm uma política que segue as recomendações, os métodos baseados em políticas visam estimar a política ideal diretamente, e o valor é um secundário, se calculado. Normalmente, uma política é parametrizada como uma rede neural. Os métodos de gradiente de política usam a descida do gradiente para estimar os parâmetros da política que maximizam a recompensa esperada. O resultado pode ser uma política estocástica onde as ações são selecionadas por amostragem ou uma política determinística [43].

2.4.6 *Ator-Crítico*

Com o passar do tempo, vários tipos de algoritmos de RL foram introduzidos e podem ser divididos em três grupos: métodos ator-somente, crítico-somente e ator-crítico, onde as palavras *ator* e *crítico* são sinônimos para a política e função de valor, respectivamente. Os métodos ator-somente geralmente funcionam com uma família parametrizada de políticas sobre as quais os procedimentos de otimização podem ser usados diretamente. O benefício de uma política parametrizada é que um espectro de ações contínuas pode ser gerado, mas os métodos de otimização usados (normalmente chamados de métodos de gradiente de política) sofrem de alta variação nas estimativas do gradiente, levando a um aprendizado lento [44].

Métodos somente críticos que usam aprendizado por diferenças temporais têm uma variância menor nas estimativas dos retornos esperados. Uma maneira direta de derivar uma política em métodos somente críticos é selecionar ações gulosas: ações para as quais a função de valor indica que o retorno esperado é o mais alto. No entanto, para fazer isso, é necessário recorrer a um procedimento de otimização em cada estado encontrado para encontrar a ação que leva a um valor ótimo. Isso pode ser computacionalmente intensivo, especialmente se o espaço de ação for contínuo. Portanto, métodos somente

críticos geralmente discretizam o espaço de ação contínua. Depois disso, a otimização sobre o espaço de ação torna-se uma questão de enumeração. Obviamente, essa abordagem prejudica a capacidade de usar ações contínuas e, portanto, de encontrar o valor ótimo [44].

Os métodos ator-crítico combinam as vantagens dos métodos somente ator e somente crítico. Enquanto o ator parametrizado traz a vantagem de computar ações contínuas sem a necessidade de procedimentos de otimização em uma função de valor, o mérito da crítica é que ele fornece ao ator um conhecimento de baixa variância do desempenho. Mais especificamente, a estimativa do retorno esperado pelo crítico permite que o ator atualize com gradientes que tenham menor variância, acelerando o processo de aprendizagem. A menor variância é trocada por um viés maior no início do aprendizado, quando as estimativas do crítico estão longe de serem precisas. Os métodos ator-crítico geralmente têm boas propriedades de convergência, em contraste com os métodos somente críticos [44].

2.4.7 *Asynchronous Advantage Actor-Critic*

Mnih *et al.* [45] propôs uma estrutura conceitualmente simples e leve para aprendizado de reforço profundo que usa descida de gradiente assíncrona para otimização de controladores de redes neurais profundas. Naquele trabalho foram apresentadas variantes assíncronas de quatro algoritmos de aprendizado por reforço padrão e os autores demonstraram que atores-aprendizes paralelos têm um efeito estabilizador no treinamento, permitindo que todos os quatro métodos treinem controladores de rede neural com sucesso. O método de melhor desempenho, uma variante assíncrona do ator-crítico denominada *Asynchronous Advantage Actor-Critic* (A3C), superou o estado da arte da época no domínio Atari enquanto treinou na metade do tempo em uma única CPU multi-núcleo em vez de uma GPU. Além disso, os autores demonstraram que o A3C é bem-sucedido em uma ampla variedade de problemas de controle motor contínuo, bem como em uma nova tarefa de navegar por labirintos 3D aleatórios usando uma entrada visual.

O A3C usa uma única rede neural profunda (DNN) para aproximar a política e a função de valor. A DNN tem duas camadas convolucionais com filtros $16 \times 8 \times 8$ com um passo de 4 e filtros $32 \times 4 \times 4$ com um passo de 2, seguido por uma camada totalmente conectada com 256 unidades. Cada camada oculta é seguida por uma função retificadora não linear. As duas saídas são uma camada *softmax* que aproxima a função de política $\pi(a_t|s_t; \theta)$ e uma camada linear para produzir uma estimativa de $V(s_t; \theta)$. Vários agentes jogam simultaneamente e otimizam o DNN por meio de descida de gradiente assíncrona. Semelhante a outros métodos assíncronos, os pesos da rede são armazenados em um servidor de parâmetros central. Os agentes calculam gradientes e enviam atualizações

para o servidor após cada $t_{max} = 5$ ações ou quando um estado terminal é atingido. Após cada atualização, o servidor central propaga novos pesos aos agentes para garantir que eles compartilhem uma política comum [46].

Capítulo 3

Conceitos Fundamentais de Gerenciamento de Tráfego Aéreo

A introdução de um novo tipo de aeronave exigirá a criação de novas regras específicas para atender à demanda que surgirá a partir do uso dos eVTOL. Entretanto, já existe um grande arcabouço de regras de tráfego aéreo existentes que não serão substituídas, mas sim adaptadas à nova realidade. Assim, é necessário conhecer tanto as instituições reguladoras quanto as regras que são aplicadas atualmente no Gerenciamento do Tráfego Aéreo (ATM) para poder compreender as mudanças que ocorrerão.

3.1 OACI

O Brasil é um dos países signatários da Organização Internacional da Aviação Civil (OACI). Abaixo será apresentada uma breve história da OACI e em seguida algumas de suas principais normas. A OACI é uma agência especializada das Nações Unidas, estabelecida na Convenção sobre Aviação Civil Internacional, ou Convenção de Chicago, que ocorreu em 1944. Esta convenção ficou conhecida como Convenção de Aviação Civil Internacional (CACI). Nesta ocasião foi produzido um documento principal, de número 7300 que foi acompanhado de, inicialmente, 18 anexos que tratam sobre áreas específicas relacionadas à aviação civil [47].

Os objetivos da OACI são desenvolver os princípios e técnicas de navegação aérea internacional e promover o planejamento e desenvolvimento do transporte aéreo internacional e facilidades de navegação aérea, encorajando o projeto de aeronaves, prevenindo o desperdício econômico causado por competição irracional, evitando discriminação entre membros, promoção de segurança e assegurando o crescimento seguro e ordenado da aviação civil internacional em todo o mundo. A OACI busca promover de forma geral o desenvolvimento de todos os aspectos da aeronáutica civil internacional. No centro da

missão da OACI está a padronização internacional da aviação civil. Cada Estado contratante se compromete a colaborar para assegurar o mais alto grau possível de uniformidade em regulamentos, padrões, procedimentos e organização em relação a aeronaves, pessoal, vias aéreas e serviços auxiliares em todos os assuntos em que tais a uniformidade facilita e melhora a navegação aérea. A constituição e a estrutura da OACI são semelhantes (mas diferentes em aspectos significativos) às Nações Unidas [48]. No Brasil, o Departamento de Controle do Espaço Aéreo (DECEA) do Comando da Aeronáutica é a organização responsável pelo controle do espaço aéreo.

3.2 Regras do Ar

As regras apresentadas abaixo são um extrato do Anexo 2 - Regras do Ar [49], que define as regras aplicáveis a todas as aeronaves em evolução no espaço aéreo de um Estado signatário da OACI.

Durante o voo, toda aeronave deverá obedecer às regras gerais e, adicionalmente, a um conjunto de regras correspondentes ao tipo de voo que está sendo realizado: as regras de voo visual (VFR) ou regras de voo por instrumentos (IFR). No voo VFR, o piloto deverá manter-se a uma certa distância de nuvens e orientar-se com base em referências no solo. O voo IFR não exige a utilização de referências no solo, porém requer instrumentos específicos na aeronave e tripulação habilitada para tal tipo de voo.

O piloto em comando de uma aeronave deverá, manipulando ou não os controles, ser responsável pela operação da aeronave de acordo com as regras do ar. Podendo desviar dessas regras apenas em circunstâncias absolutamente necessárias no interesse da segurança. O piloto em comando de uma aeronave terá autoridade final quanto à disposição da aeronave enquanto estiver em comando.

Exceto quando necessário para decolagem ou pouso, ou exceto com permissão da autoridade competente, as aeronaves não devem sobrevoar as áreas congestionadas de cidades, vilas ou assentamentos ou sobre uma reunião de pessoas ao ar livre, a menos que a uma altura tal que permitirá, no caso de surgir uma emergência, um pouso a ser feito sem perigo indevido para pessoas ou bens na superfície.

Uma aeronave não deve ser operada tão próxima de outra aeronave que crie um risco de colisão. O sistema embarcado para evitar colisões (ACAS) que mostra ao piloto de uma aeronave a posição relativa de outra aeronave voando nas proximidades, sempre que o intruso estiver com o transponder em funcionamento. Cabe inteiramente ao piloto em comando de uma aeronave a responsabilidade de realizar manobras para evitar colisões com outras aeronaves a partir dos avisos de resolução fornecidos pelo ACAS.

Uma aeronave em voo, ou em operação no solo ou na água, deve dar passagem à aeronave em pouso ou em fase final de aproximação para pouso. Quando duas ou mais aeronaves mais pesadas que o ar estiverem se aproximando de um aeródromo para fins de pouso, as aeronaves do nível superior devem ceder passagem às aeronaves do nível inferior, mas estas últimas não devem se beneficiar desta regra cortar na frente de outra que esteja em fase final de aproximação para pouso, ou ultrapassar aquela aeronave. No entanto, aeronaves mais pesadas que o ar com propulsão mecânica devem dar lugar a planadores.

As informações relativas a um voo pretendido ou parte de um voo, a serem fornecidas aos órgãos dos serviços de tráfego aéreo, devem estar na forma de um plano de voo.

Um plano de voo deve ser submetido, antes da partida, a uma sala de informações aeronáuticas ou, durante o voo, transmitido à unidade apropriada dos serviços de tráfego aéreo ou à estação de rádio de controle ar-solo, a menos que tenham sido tomadas providências para a apresentação de planos de voo repetitivos. Entre as informações presentes em um plano de voo estão:

- Identificação da aeronave;
- Regras de voo e tipo de voo;
- Número e tipo de aeronave e categoria de esteira de turbulência;
- Equipamento;
- Aeródromo de partida;
- Tempo estimado de remoção dos calços;
- Velocidade de cruzeiro;
- Nível de cruzeiro;
- Rota a seguir;
- Aeródromo de destino e tempo decorrido total estimado;
- Aeródromos alternativos;
- Autonomia de combustível;
- Número total de pessoas a bordo; e
- Equipamentos de emergência e sobrevivência.

Uma autorização de controle de tráfego aéreo deve ser obtida antes de operar um voo controlado, ou uma parte de um voo como um voo controlado. Tal autorização deverá ser

solicitada mediante a apresentação de um plano de voo a um órgão de controle de tráfego aéreo.

Uma aeronave deve aderir ao plano de voo atual ou à parte aplicável de um plano de voo atual submetido para um voo controlado, a menos que uma solicitação de alteração tenha sido feita e a autorização obtida da unidade de controle de tráfego aéreo apropriada, ou a menos que surja uma situação de emergência que exija ação imediata da aeronave, caso em que, assim que as circunstâncias permitirem, após o exercício dessa ação emergencial, a unidade de serviços de tráfego aéreo apropriada deverá ser notificada da ação tomada e que esta ação foi tomada devido a uma emergência.

A menos que autorizado de outra forma pela autoridade de Serviços de Tráfego Aéreo (ATS) apropriada, ou ordenado pela unidade de ATC apropriada, os voos controlados deverão, na medida do possível: a) quando em uma rota ATS estabelecida, operar ao longo da linha central definida dessa rota; ou b) quando em qualquer outra rota, operar diretamente entre os auxílios à navegação e/ou pontos que definem aquela rota.

Uma aeronave operada como um voo controlado deve manter comunicação de voz ar-terra contínua no canal de comunicação apropriado e estabelecer comunicação bidirecional conforme necessário com o órgão de controle de tráfego aéreo apropriado, exceto conforme prescrito pela autoridade ATS apropriada em relação a aeronaves que fazem parte do tráfego de aeródromo em um aeródromo controlado.

3.3 Serviços de Tráfego Aéreo

As definições internacionais sobre os serviços de tráfego aéreo são apresentadas no Anexo 11 da OACI [50]. A norma do DECEA por meio da qual são divulgadas a prestação destes serviços no território nacional é a ICA 100-37 [51] que também é denominada *Serviços de Tráfego Aéreo*.

O espaço aéreo é dividido em porções conforme a denominação a seguir:

Regiões de informação de voo As porções do espaço aéreo onde for determinado que o serviço de informações de voo e o serviço de alerta estarão disponíveis serão designadas como regiões de informações de voo. A divisão do território do Brasil em regiões de informação de voo é apresentada na Figura 3.1. A região que aparece na cor amarela corresponde ao Centro Regional de Controle do Espaço Aéreo Sudeste (CRCEA-SE) que possui atribuições diferentes de uma FIR, mas está contido em grande parte na FIR Curitiba.

Áreas de controle e zonas de controle As porções do espaço aéreo onde for determinado que o serviço de controle de tráfego aéreo será fornecido para voos IFR devem ser designadas como áreas de controle ou zonas de controle. Quando designadas

dentro de uma região de informação de voo, as áreas de controle e as zonas de controle devem fazer parte dessa região de informação de voo.

Aeródromos controlados Os aeródromos onde for determinado que o serviço de controle de tráfego aéreo será fornecido ao tráfego do aeródromo serão designados como aeródromos controlados.



Figura 3.1: Regiões de informação de voo do Brasil.

Os espaços aéreos ATS serão classificados e designados de acordo com o tipo de voo aceito, a separação provida entre as aeronaves, o serviço prestado e os requisitos para utilização conforme Figura 3.2:

As áreas de controle, incluindo, aerovias e áreas de controle terminal, devem ser delimitadas de modo a abranger espaço aéreo suficiente para conter as trajetórias de voo dos voos IFR ou partes dos quais se deseja fornecer às partes aplicáveis do serviço de controle de tráfego aéreo, tendo em conta as capacidades dos auxílios à navegação normalmente utilizados naquela área.

Os limites laterais das zonas de controle devem abranger pelo menos aquelas porções do espaço aéreo, que não estão dentro das áreas de controle, contendo as trajetórias de voos

CLASSE	TIPO DE VOO	SEPARAÇÃO PROVIDA	SERVIÇO PRESTADO	REQUISITOS PARA UTILIZAÇÃO	
				Requisitos de radiocomunicação	Sujeito a uma autorização ATC
A	IFR	A todas aeronaves	Serviço de controle de tráfego aéreo	Bilateral contínua	Sim
B	IFR / VFR	A todas aeronaves	Serviço de controle de tráfego aéreo	Bilateral contínua	Sim
C	IFR	IFR de IFR IFR de VFR	Serviço de controle de tráfego aéreo	Bilateral contínua	Sim
	VFR	VFR de IFR	1) Serviço de controle de tráfego aéreo para separação de IFR; e 2) Informação de tráfego VFR/VFR e aviso para evitar tráfego, quando solicitado pelo piloto.		
D	IFR	IFR de IFR	Serviço de controle de tráfego aéreo, incluindo informação de tráfego sobre voos VFR (e aviso para evitar tráfego, quando solicitado pelo piloto)	Bilateral contínua	Sim
	VFR	Não aplicável	Serviço de controle de tráfego aéreo, incluindo informação de tráfego sobre voos IFR/VFR e VFR/VFR (e aviso para evitar tráfego, quando solicitado pelo piloto)		
E	IFR	IFR de IFR	Serviço de controle de tráfego aéreo. Informações de tráfego sobre voos VFR sempre que praticável.	Bilateral contínua	Sim
	VFR	Não aplicável	Informações de tráfego sempre que praticável.	Não	Não

Figura 3.2: Classificação dos espaços aéreos ATS.

IFR chegando e partindo de aeródromos para serem usados em condições meteorológicas por instrumentos.

Os serviços de tráfego aéreo compreenderão três serviços denominados serviço de controle de tráfego aéreo, serviço de informação de voo e serviço de alerta. Os serviços de controle de tráfego subdividem-se em:

- Serviço de controle de área: a prestação do serviço de controle de tráfego aéreo para voos controlados.
- Serviço de controle de aproximação: a prestação do serviço de controle de tráfego aéreo nas partes de voos controlados associadas à chegada ou partida.
- Serviço de controle de aeródromo: a prestação do serviço de controle de tráfego aéreo para o tráfego de aeródromo.

Quando for determinado que os serviços de tráfego aéreo serão prestados em determinadas porções do espaço aéreo ou em determinados aeródromos, essas porções do espaço

aéreo ou aeródromos deverão ser designadas em relação a quais tipos de serviços de tráfego aéreo.

O serviço de controle de área é prestado por órgãos denominados Centro de Controle de Área (ACC). Por sua vez, o serviço de controle de aeródromo é prestado por uma torre de controle de aeródromo (TWR). O Serviço de controle de aproximação poderá ser prestado 1) por uma TWR ou por um ACC quando for necessário ou desejável combinar sob a responsabilidade de uma unidade as funções do serviço de controle de aproximação com as do serviço de controle de aeródromo ou do serviço de controle de área; ou 2) por uma unidade de controle de aproximação (APP) quando for necessário ou desejável estabelecer uma unidade separada.

Os objetivos dos serviços de tráfego aéreo são:

- Evitar colisões entre aeronaves;
- Prevenir colisões entre aeronaves na área de manobras dos aeródromos
- Agilizar e manter o fluxo ordenado do tráfego aéreo
- Fornecer informações úteis para a condução segura e eficiente dos voos
- Notificar as organizações apropriadas a respeito de aeronaves que necessitem de ajuda de busca e salvamento e auxiliar essas organizações conforme necessário.

Quando as rotas ATS são estabelecidas, um espaço aéreo protegido ao longo de cada rota ATS e um espaçamento seguro entre as rotas ATS adjacentes devem ser fornecidos. As rotas ATS devem ser identificadas por designadores.

Cada Estado poderá determinar, de acordo com as disposições no Anexo 11 e para os territórios sob sua jurisdição, as porções do espaço aéreo e os aeródromos onde serão prestados serviços de tráfego aéreo. Posteriormente, providenciarão para que tais serviços sejam estabelecidos e prestados de acordo com as disposições deste Anexo, exceto que, por acordo mútuo, um Estado poderá delegar a outro Estado a responsabilidade de estabelecer e prestação de serviços de tráfego aéreo em regiões de informação de voo, áreas de controle ou zonas de controle que se estendam sobre os territórios daquelas.

3.3.1 Separação entre Aeronaves

Existem diversos métodos estabelecidos para garantir uma separação longitudinal segura entre aeronaves em rota. As distâncias entre as aeronaves envolvidas podem chegar a valores de até 15 minutos de voo para a separação por tempo ou até 150 quilômetros para a separação baseada em distância por Navegação de Área (RNAV). Uma vez que o serviço de vigilância ATS está disponível em praticamente todo o espaço aéreo Brasileiro,

a separação empregada na prática é a de 18 km para os espaços aéreos servidos apenas por radar de rota ou 9,2 km para as demais regiões [51].

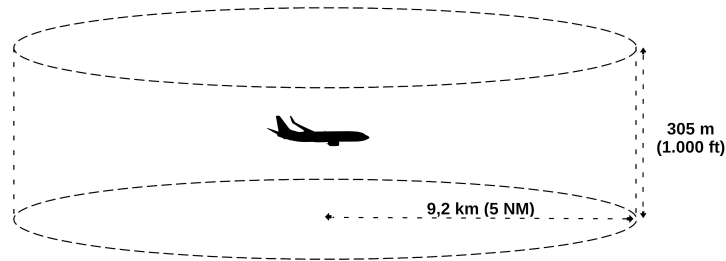


Figura 3.3: Separação horizontal e vertical.

A separação horizontal é de 300 metros ou 1.000 pés até o nível de voo 410, ou seja, até uma altitude de 12,5 km acima do nível do mar. Essa separação pode ser aplicada no espaço aéreo onde a Separação Vertical Mínima Reduzida (RVSM) é empregada [51]. A composição da separação longitudinal e vertical cria uma região de proteção ao redor da aeronave conforme ilustração da Figura 3.3.

3.3.2 Resolução de Conflitos

Caso a projeção das rotas de duas ou mais aeronaves indique que em algum momento a separação ficará abaixo daquela aplicada no espaço aéreo em questão, isso indica um conflito, conforme ilustração Figura 3.4.

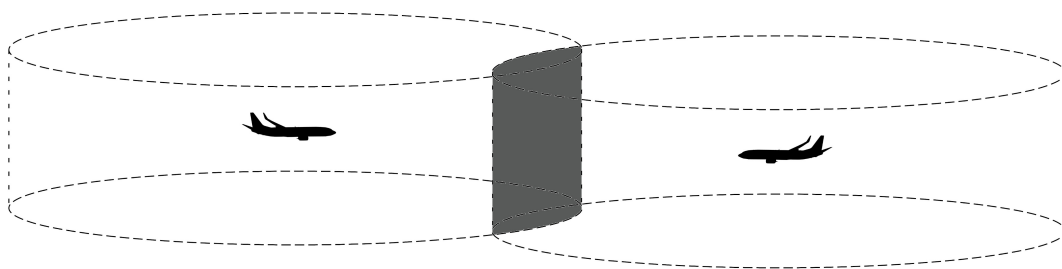


Figura 3.4: Conflito com perda de separação horizontal.

A resolução de conflitos é dividida em duas categorias, de acordo com o momento em que a manobra para a resolução do conflito é realizada. Caso a manobra seja realizada em até 20 minutos antes do momento previsto para conflito, considera-se a uma resolução estratégica. Se a resolução ocorrer com uma antecedência menor, geralmente de 3 a 5 minutos, essa manobra é chamada de resolução tática [52].

Um conflito entre duas aeronaves pode ser descrito em termos de informações sobre as posições relativas, direções ou proas de cada uma das aeronaves e distância. Esse conjunto de informações é denominado geometria do conflito [53]. A figura 3.5 ilustra como a proa relativa é importante para permitir que um sistema analise e busque uma resolução para o conflito. Esta informação indica, de forma indireta, o ponto de cruzamento entre as duas aeronaves. Portanto, pode ser utilizada para determinar a manobra mais eficiente para obter a separação necessária.

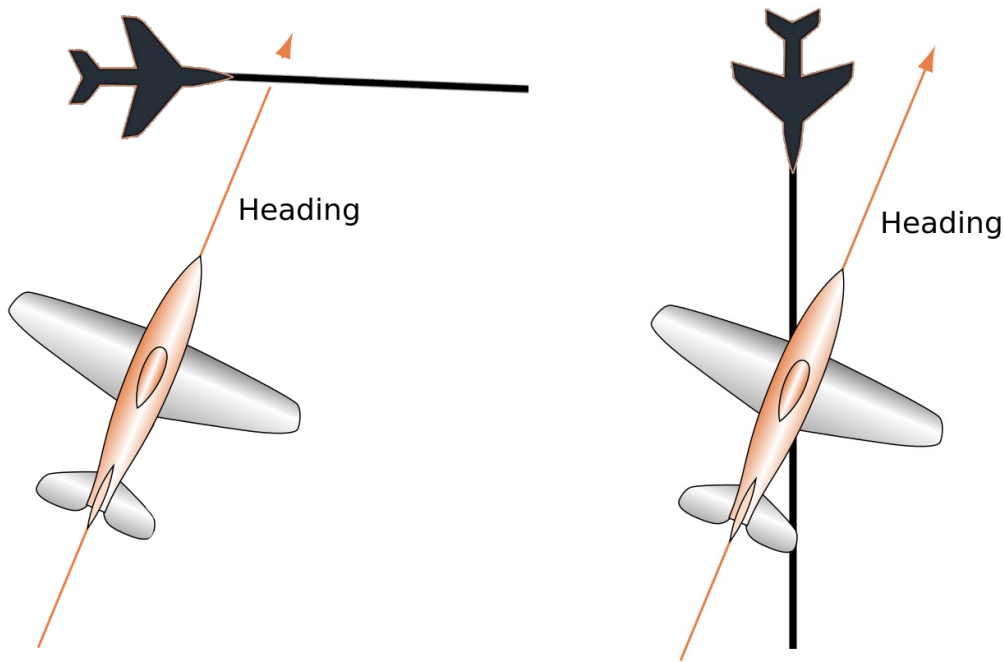


Figura 3.5: Proa relativa entre duas aeronaves em conflito.

3.3.3 Voo Livre

O voo livre, também conhecido como espaço aéreo de rotas livres (FRA), é uma condição de navegação e planejamento de rota nas quais os operadores têm a liberdade de selecionar seu caminho e velocidade em tempo real. As regras aplicáveis ao UAM ainda estão por ser definidas, porém espera-se que o voo livre seja aplicado a alguns trechos do espaço aéreo que será utilizado pelos eVTOL.

As restrições de tráfego aéreo são impostas apenas para garantir a separação, evitar exceder a capacidade do aeroporto, evitar voos não autorizados através de porções restritas do espaço aéreo e garantir a segurança do voo. As restrições são limitadas em extensão e duração para corrigir o problema identificado. Qualquer atividade que remova restrições representa um movimento em direção ao voo livre. Isso sugere que cada usuário recebe

flexibilidade máxima e separação segura garantida. O objetivo não é apenas otimizar o sistema, mas abrir o sistema para que cada usuário possa fazer a sua própria otimização. No voo livre, cada aeronave deve ter a liberdade de voar em uma trajetória otimizada, fazendo pleno uso dos sistemas de gerenciamento de voo a bordo.

[10]

Nessa modalidade, as aeronaves são totalmente responsáveis por sua própria separação de outro tráfego, elas são livres para decidir sobre sua rota ideal, ou seja, rota direta, em vez de seguir a rota recebida de um órgão central. Uma vez que a tecnologia ADS-B ou outra semelhante estiver desenvolvida ao ponto de garantir maior confiabilidade e desempenho, um sistema distribuído de resolução de conflitos pode garantir com segurança a separação entre aeronaves [54].

3.4 Cenário Previsto para UAM

Apesar do cenário atual de regras e suas aplicações para o gerenciamento de tráfego aéreo da aviação convencional ser amplamente conhecido e pesquisado, tal situação se inverte quando o assunto é a mobilidade aérea urbana. Um dos principais fatores que contribui para as indefinições são as próprias aeronaves eVTOL que ainda estão em fase de desenvolvimento e não podem ser utilizadas em testes de larga escala. Ainda há um número bastante pequeno de publicações que contenham regras relacionadas ao UAM. Porém já existem diversos Conceitos Operacionais (ConOps) antevendo os cenários de aplicação de aeronaves específicas em algumas capitais ao redor do mundo.

Abaixo será um dos trabalhos já existentes nos quais o cenário da UAM é antecipado. Este possível cenário de aplicação será necessário para nortear a revisão bibliográfica bem como os demais capítulos deste trabalho.

3.4.1 Níveis de Maturidade

Considerando que a adoção das aeronaves eVTOL ocorrerá de forma gradual e que adaptações serão feitas durante todo o processo, a Administração Nacional de Aeronáutica e Espaço (NASA) desenvolveu o conceito de níveis de maturidade do UAM (UML) para avaliar como será o desenvolvimento dos três principais componentes envolvidos: aeronaves, espaço aéreo e comunidade [55]. Abaixo serão descritas as principais características de cada um dos seis níveis de maturidade presentes no trabalho mencionado, visto que as etapas de implantação do UAM serão utilizadas em outros pontos deste texto.

UML-1

O primeiro nível de maturidade, (UML-1) refere-se a um período pré-operacional, sendo caracterizado, da perspectiva do desenvolvimento de aeronaves, pela existência de um projeto de certificação de aeronaves em estágio avançado. Os fabricantes e a autoridade certificadora devem concordar em relação a um plano de certificação específico para cada projeto de aeronave após a sua conclusão. Tal plano define as análises e os testes, em solo e voo, que serão aplicados para demonstrar a conformidade com os regulamentos de certificação aplicáveis. As empresas que trabalham para alcançar o UML-1 para aeronaves não pilotadas enfrentarão significativamente mais incertezas do que os desenvolvedores de aeronaves pilotadas convencionalmente, dada a estrutura regulatória ainda incipiente que abrange operações de sistema de aeronaves não tripuladas. Nessa fase, é possível que os fabricantes executem demonstrações em possíveis mercados de adoção precoce para fins de avaliação da integração da comunidade e aceitação de questões como ruído, adequação de instalações de solo e integração nas operações locais do espaço aéreo.

UML-2

O UML-2 é caracterizado por tráfego de baixa densidade, operações de baixa complexidade e dependência da automação assistiva. Este nível provavelmente estará relacionado com o início da capacidade de transporte comercial de passageiros nas operações UAM. As operações nesta fase ocorrem em um número limitado de localidades onde há ambiente favorável e apoio político local para a introdução do serviço UAM, sendo introduzidos de forma controlada e gradual. A certificação dos pilotos exigirá ao menos uma licença de piloto comercial. Dadas as limitações iniciais do número de aeronaves e pilotos disponíveis, espera-se que existam preocupações mínimas de segurança relacionadas a densidades operacionais. Considerando o atributo de dependência da automação, os sistemas de controle de voo relativamente sofisticados e de alta autoridade provavelmente serão necessários para ajudar os pilotos a alcançar os níveis necessários de segurança, desempenho e robustez.

O design do espaço aéreo e os requisitos de integração operacional são relativamente simples no UML-2, com apenas modificações mínimas para os procedimentos atuais necessários para acomodar operações de UAM de baixo volume. A integração eficiente das operações da UAM no espaço aéreo controlado em torno dos principais aeroportos comerciais provavelmente já deverá ser levada em consideração. O acesso ao espaço aéreo para as operações não tripuladas envolve barreiras significativas com estratégias de solução ainda incertas. É provável que as operações iniciais do espaço aéreo para operações ocorram em espaço aéreo relativamente pouco utilizado, longe das áreas urbanas. As

operações iniciais provavelmente serão restritas a menos áreas sensíveis ao ruído, como distritos comerciais e bairros próximos a aeroportos comerciais.

UML-3

O UML-3 é caracterizado por baixa densidade de tráfego, operações de complexidade média e dependência de automação de segurança abrangente. Será um período de transição em que as tecnologias necessárias para apoiar operações mais avançadas e escaláveis sejam introduzidas e avaliadas operacionalmente, mas com crédito operacional limitado concedido até que dados operacionais significativos e validação das novas tecnologias sejam obtidas. Portanto, espera-se que as características definidoras do UML-3 sejam o crescimento operacional inicial além do UML-2, mas com escalabilidade e implantação limitadas fora das áreas de adoção precoce até que as alterações regulatórias permitam o crédito operacional para a integração das principais tecnologias e capacidades. As principais tecnologias introduzidas durante a UML-3 estão relacionadas à escalabilidade dos recursos humanos como, por exemplo, pilotos e controladores; escalabilidade das operações do espaço aéreo; estabelecimento de atingir limites de ruído para aceitação da comunidade; e a capacidade de operar em condições de voo por instrumentos.

Depois que o desempenho do veículo, a segurança e as características de ruído aceitáveis para uma comunidade forem demonstradas, as cidades poderão considerar a aprovação das vertiportos iniciais em áreas altamente desenvolvidas. Para que a UAM forneça economias de tempo significativas em relação às alternativas de superfície, a capacidade de implantar vertiportos próximos ou dentro dos centros populacionais é um avanço importante além do UML-2.

Embora o número de aeronaves em voo simultâneo sobre uma determinada cidade seja relativamente limitado, a densidade das operações ao redor de vertiportos importantes poderá ser eventualmente alta. O gerenciamento eficiente de decolagem e janelas de tempo para os pousos serão de vital importância para a segurança, experiência de passageiros, minimização de atrasos em voo e maximização da taxa de transferência do sistema. Antes da decolagem, eventuais conflitos entre os voos serão resolvidos de forma estratégica. Espera-se que os voos sejam realizados sob o VFR no UML-3, a responsabilidade final de "ver e evitar" outro tráfego permanece com os pilotos de aeronaves e os Provedores de Serviços UAM (PSU) nesta fase fornecem serviços de apoio.

A interação com o controle de tráfego aéreo tradicional nas classes de espaço aéreo gerenciado pelo controlador usado pelas operações da UAM durante o UML-3 (por exemplo, classe B, C, D) ainda precisará cumprir os requisitos regulatórios atuais. Para minimizar a carga de trabalho adicional em controladores e pilotos, espera-se que grande parte da comunicação de rotina seja tratada de forma transparente através da conectividade digital.

A comunicação de voz padrão ainda precisará estar disponível no espaço aéreo gerenciado pelo controlador para interações não rotineiras. Espera-se que a comunidade trabalhe para desenvolver equipamentos e procedimentos do espaço aéreo, incluindo regras de voo e procedimentos de voo.

UML-4

A UML-4 é caracterizada por níveis de tráfego médio, operações de complexidade média e dependência de automação colaborativa e responsável. Prevê-se que o UML-4 seja ativado após várias alterações regulatórias importantes que aumentam significativamente a confiabilidade do transporte da UAM e sua escalabilidade. Essas mudanças regulatórias estenderiam as operações da UAM para condições meteorológicas por instrumentos e reduziriam as habilidades especializadas e o treinamento associado necessário aos pilotos e controladores humanos, à medida que sistemas automatizados de alta garantia são confiáveis para executar funções selecionadas e críticas de segurança. Na UML-4, espera-se que a UAM seja uma realidade cotidiana em muitas áreas metropolitanas. Além disso, espera-se que as economias crescentes de escala possam tornar a UAM acessível e atraente para uma porcentagem significativa do público para viagens entre locais de alta densidade de passageiros como, por exemplo, aeroportos comerciais e áreas comerciais.

A principal diferença entre aeronaves no UML-4 em relação ao nível anterior é que a automação abrangente de voo e gerenciamento de contingências introduzida no UML-3 foi refinada e validada operacionalmente. Além disso, os requisitos de certificação e treinamento para pilotos são atualizados para refletir as capacidades e a confiabilidade dessa tecnologia. Embora essas revisões possam assumir muitas formas, uma das estratégias gerais que estão sendo perseguidas pela indústria é alavancar os recursos de automação no ar para permitir a aprovação de que o piloto esteja fora do veículo. Essa abordagem, muitas vezes chamada de operações de supervisão remota (RSO), tem a vantagem óbvia de remover o piloto e a estação do piloto associado da carga útil nominal do veículo, que é bastante significativa para uma aeronave para aproximadamente quatro assentos.

É provável que o sistema de espaço aéreo primário dos avanços exigidos para a UML-4 seja facilitado por mudanças regulatórias e políticas, permitindo um aumento de crédito operacional para tecnologias implantadas no UML-3. Essas alterações seriam direcionadas para ativar os recursos de gerenciamento de operações do espaço aéreo da UAM implementados e validados no VMC a serem utilizados no IMC no UML-4.

Neste nível, a UAM cresce além do número relativamente limitado de cidades de adiantes iniciais e se torna prático em uma ampla gama de locais. Essa expansão é ativada pela combinação de poder operar com clima de baixa visibilidade; capacidade de localizar vertiportos de alta capacidade em áreas densamente povoadas, garantindo a segurança dos

vizinhos; níveis de ruído aceitáveis para comunidades; e crescentes economias de escala que tornam a capacidade amplamente acessível aos membros da comunidade.

UML-5

A UML-5 é caracterizada por operações de alta densidade e complexidade, dependendo de redes automatizadas altamente integradas. Comparado ao UML-4, o sistema possui uma ordem de magnitude a mais de aeronaves e vertiportos em uma determinada área metropolitana. Para comparação, o número total de aeronaves da aviação geral e aviação comercial em voo sobre o território continental dos EUA durante os períodos de pico hoje é da ordem de 5.000. Nesse caso, a estimativa de voos simultâneos seria em torno de 50.000 aeronaves.

Embora o nível projetado de atividade aérea local seja muito alto para os padrões atuais, deve-se reconhecer que, mesmo nessa escala, o número de aeronaves ainda será relativamente limitado em comparação com as frotas típicas de carros e caminhões leves na área. Devido ao aumento da velocidade, dos fatores de carga e da utilização, a produtividade de aeronaves individuais será muitas vezes maior que a dos carros individuais. Com base nessas projeções, no UML-5, espera-se que a distância total de passageiros percorrida por aeronaves UAM em uma área metropolitana seja da ordem de 4-5% da frota pessoal de veículos terrestres. A motivação para essa comparação é destacar que, nos níveis de tráfego UML-5, que são muito altos pelos padrões atuais de aviação, alta produtividade, as operações de frota são essenciais para o sistema fornecer níveis significativos de transporte de passageiros em comparação com o sistema de transporte terrestre atual .

UML-6

O alcance do UML-6 é caracterizado por operações e instalações onipresentes da UAM em uma área metropolitana. Uma definição mais operacionalmente tangível de UML-6 é quando a construção da UAM é essencialmente completa e o crescimento adicional do sistema corresponde ao crescimento da população dentro de uma área metropolitana. Obviamente, há um alto grau de incerteza sobre como esse sistema maduro pode ser, pois provavelmente ele só se tornará uma realidade em várias décadas e várias gerações de tecnologia no futuro. No UML-5, com centenas de portos distribuídos pela região, a maioria das pessoas viverá a poucos quilômetros de um vertiporto. Como a maioria das viagens exigirá segmentos terrestres, principalmente de carro, para chegar ou sair dos vertiportos, a capacidade seria principalmente vantajosa para viagens mais longas, com os carros ainda sendo a escolha predominante para viagens mais curtas. Para que o UAM obtenha uma utilização aumentada além do UML-5, é provável que ele precise

operar em modo porta-a-porta (por exemplo, entre vertiportos da vizinhança) para um número significativo de viagens em potencial, tornando-o vantajoso em comparação com o transporte terrestre para viagens abaixo de 20 quilômetros.

3.4.2 Resolução de Conflitos em UAM

Assim como no ATM, a resolução de conflitos no UAM se dará em várias fases, permitindo que um conflito seja detectado e resolvido com a maior antecedência possível [56]. A Figura 3.6 mostra um resumo das fases com as ações que podem ser tomadas em cada etapa, incluindo a janela temporal onde as medidas podem ser aplicadas [54].

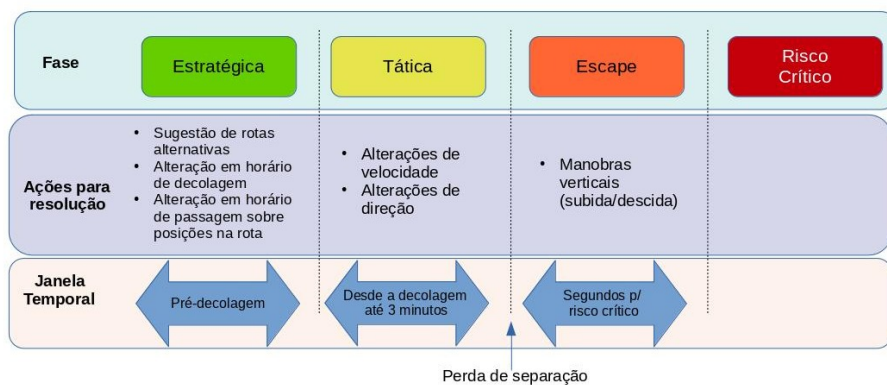


Figura 3.6: Fases da resolução de conflito no UAM. Adaptado de Ribeiro *et al.* [54]

A fase estratégica da resolução de conflitos ocorre normalmente antes da decolagem. Nesta etapa, um pré-processamento dos planos de voo detecta os conflitos que poderão ocorrer em uma determinada porção do espaço aéreo e oferece as opções de medidas que podem ser tomadas para a resolução de cada um dos conflitos. Por se tratar de modificações no plano de voo, poderá ser aplicado o conceito da Tomada de Decisão Colaborativa (CDM), no qual todos os envolvidos são chamados a participar da decisão de modo a obter uma solução final que seja satisfatória para o conjunto de participantes [57]. Entre as medidas que podem ser aplicadas na fase estratégica estão a alteração do horário de decolagem, sugestão de rotas alternativas e também alterações no horário previsto de passagem sobre determinados pontos da rota.

A fase tática de resolução de conflitos ocorre durante o voo, geralmente com antecedência entre 5 a 3 minutos do momento previsto para o CPA, mas pode ocorrer a qualquer momento a partir da decolagem. Neste momento ainda é possível tomar medidas que sejam capazes de garantir uma separação adequada. Alterações nas velocidades e/ou nas direções das aeronaves envolvidas são as medidas mais indicadas.

Caso nenhuma medida de resolução de conflitos tenha sido aplicada anteriormente, chega o ponto em que ocorre a perda de separação, ou seja, a distância entre as aeronaves atinge um valor abaixo dos mínimos estabelecidos para aquele espaço aéreo. Neste momento ainda é possível evitar uma colisão, desde que as medidas adequadas sejam aplicadas. Assim como na aviação geral em que o Sistema Embarcado para Evitar Colisões (ACAS) realiza manobras verticais como última medida para evitar uma colisão [58], um módulo do ACAS com o qual os eVTOL serão equipados tomará as medidas necessárias para evitar o risco crítico.

Capítulo 4

Revisão Bibliográfica

O campo de pesquisa no qual este trabalho se encontra é destinado à Detecção e Resolução de Conflitos (CDR) em tráfego aéreo. Impulsionado pelo crescimento praticamente constante do tráfego aéreo em todo o mundo nas últimas décadas, bem como pela previsão de que a demanda continue a aumentar, este campo conta com significativos avanços que foram revisados por Ribeiro et al. [54] em um trabalho dedicado a métodos de CDR incluindo mais de uma centena de artigos. Recentemente, a adoção em larga escala dos veículos aéreos não tripulados (UAV), bem como a previsão de início de utilização dos eVTOL fez com que o interesse por esse campo de pesquisa fosse renovado.

Cabe ressaltar que os trabalhos abordados aqui não são diretamente comparáveis entre si. Mesmo a comparação com o presente trabalho apresenta limitações causadas pela diferença de abordagem e pela configuração do problema que, em muitos casos, é significativamente diferente.

Esta revisão da literatura será dividida em duas partes: a primeira será dedicada a uma visão geral sobre os métodos de CDR enquanto a segunda parte será dedicada aos métodos de DRL que tem apresentado melhores resultados na resolução dos problemas de CDR.

4.1 Métodos de Detecção e Resolução de Conflitos

Além da revisão de diversos trabalhos, Ribeiro et al. [54] também desenvolveu uma taxonomia para classificação dos métodos considerando várias características dos sistemas avaliados. Este trabalho não irá aplicar todos os elementos de classificação, porém serão analisados os elementos daquela taxonomia que se aplicam ao cenário esperado para a utilização dos eVTOL na mobilidade aérea urbana.

Em relação à resolução de conflitos, a principal divisão está relacionada ao gerenciamento da separação entre as aeronaves envolvidas em um conflito, ou seja, qual entidade

determinará a realização de manobras para a solução de um conflito. As duas categorias nesta divisão são: controle centralizado e descentralizado. No controle centralizado, um único órgão é responsável por determinar ajustes de velocidade, nível de voo ou proa a todas as aeronaves envolvidas. O controle descentralizado, por sua vez, implica que cada uma das aeronaves tomará suas próprias decisões em relação às manobras. Mesmo no controle descentralizado, existe a possibilidade de comunicação entre as duas ou mais aeronaves envolvidas no conflito para encontrar a melhor solução.

Este trabalho considera o cenário onde um grande número de aeronaves eVTOL operarão de forma simultânea, a partir do nível de maturidade 4 (UML-4) descrito em 3.4.1. Essa demanda aumenta a possibilidade de erros de comunicação tanto na transmissão quanto na recepção, bem como os possíveis erros na execução de manobras evasivas. Por isso, o trabalho de Xue [59] indica que o melhor resultado será obtido com a resolução de conflitos descentralizada. Além disso, a alta demanda torna computacionalmente caro um controle centralizado. A resolução de conflitos em nível estratégico poderá ser executada por um prestador de serviços centralizado durante o processamento dos planos de voo, porém será imprescindível um componente embarcado que seja capaz de resolver os conflitos que surjam em nível tático.

Grande parte dos trabalhos que se dedicam a estudar o problema de CDR consideram o controle centralizado [54]. Porém o cenário do UAM descrito acima possui características bastante específicas para as quais não é descartada a existência de um controle centralizado, mas tais características fazem com que seja necessário o desenvolvimento de soluções complementares que não podem se beneficiar inteiramente dos resultados já obtidos em trabalhos que aplicam os conceitos de Operações Baseadas em Trajetórias (TBO).

O método denominado Potencial de Voltagem Modificado (MVP) foi criado por Hoekstra [60] *et al.* com o objetivo de realizar manobras para evitar colisões em um espaço aéreo onde é aplicado o conceito de voo livre. Nessa abordagem, as aeronaves são consideradas como partículas carregadas eletricamente que se repelem mutuamente. Essa força de repulsão é convertida em um deslocamento da posição prevista no ponto de aproximação mais próximo (CPA), de forma que a distância mínima seja igual à separação mínima exigida entre as aeronaves. Tal deslocamento resulta em um novo rumo e velocidade recomendados, na direção que aumenta o CPA previsto conforme ilustração na figura 4.1. A escolha dessa direção para cada resolução exige que exista uma coordenação entre as duas aeronaves envolvidas no conflito.

No caso de conflitos de múltiplas aeronaves, os vetores de resolução são somados para cada par de conflitos. Este método tem a vantagem da simplicidade, pois os cálculos resultantes são computacionalmente leves, e a representação geométrica permite que outras

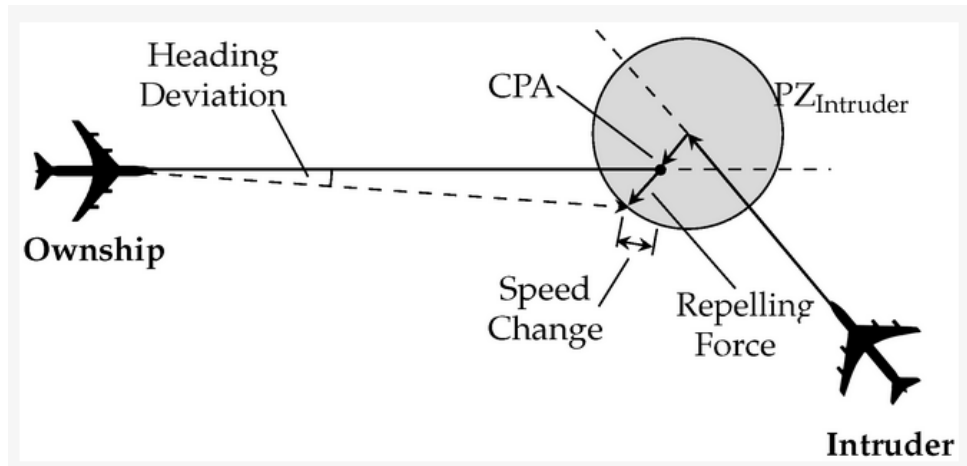


Figura 4.1: Ilustração do método Potencial de Voltagem Modificado (MVP) [60].

restrições possíveis sejam levadas em consideração facilmente. Por outro lado, como as resoluções são baseadas apenas na geometria do conflito, elas podem se opor à direção de voo desejada, levando a um aumento na distância percorrida e conseqüente aumento no consumo de combustível.

Balasooriyan [61] estudou métodos de resolução de conflitos descentralizados e implicitamente coordenados. Para compreender o comportamento dos métodos de resolução de conflitos em cenários de conflito com várias aeronaves foi utilizado o Diagrama de Espaços de Solução (SSD) como um método automatizado e horizontal de resolução de conflitos. O SSD, construído a partir de obstáculos de velocidade, indica as combinações de rumos e velocidades que levarão a trajetórias com ou sem conflitos. Além disso, o SSD abrange todas as aeronaves próximas simultaneamente, o que o torna adequado para uso na resolução de conflitos entre várias aeronaves. Foi feita uma comparação com o método MVP. Essa comparação é feita usando simulações em tempo acelerado, onde oito diferentes conjuntos de regras de coordenação foram empregados em paralelo com o SSD. As simulações mostraram que o MVP supera o método de resolução de conflitos proposto em termos de segurança, estabilidade e eficiência conforme resultados apresentados na figura 4.2.

Os efeitos da dinâmica de manobra real na separação mínima resultante foram analisados por Paielli [62], no trabalho que apresentou um algoritmo numérico simples para compensar esses efeitos. Foram considerados apenas voos nivelados, portanto toda a dinâmica de desvios ocorre no plano horizontal. As mudanças de velocidade são modeladas como períodos de aceleração ou desaceleração constante ao longo da rota, e as mudanças de direção são modeladas como curvas de raio constante. A abordagem apresentada usa modelos cinemáticos simples, nos quais é considerada uma massa pontual restrita, para obter precisão semelhante àquela dos modelos dinâmicos, mantendo grande parte da

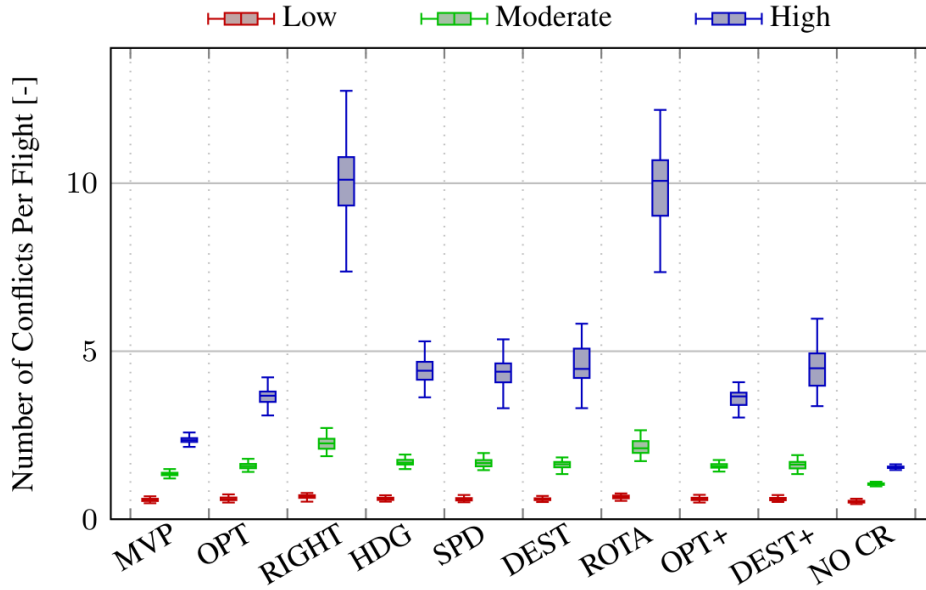


Figura 4.2: Diagrama de Espaços de Solução (SSD) [61].

simplicidade dos modelos de manobra instantânea (figura 4.3).

Algoritmos numéricos foram apresentados para determinar as mudanças de velocidade ou direção necessárias para resolver conflitos, compensando implicitamente os efeitos da dinâmica de manobra. Uma grande vantagem operacional desses algoritmos é que eles determinam imediatamente se uma determinada manobra tem tempo suficiente para resolver o conflito. Para a resolução de conflitos táticos com dinâmica de manobra realista, os algoritmos mostraram-se mais precisos do que as equações padrão baseadas em mudanças instantâneas de rumo ou de velocidade. A melhoria da precisão não é particularmente significativa para manobras de mudanças de direção, a menos que o conflito seja iminente e o ângulo de trajetória seja bastante pequeno.[62].

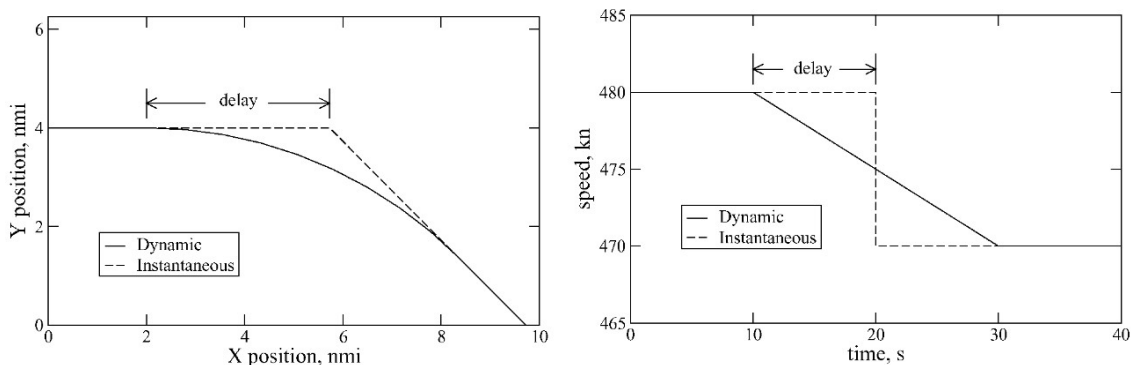


Figura 4.3: Mudança gradual de direção e velocidade [62].

Van Dam [63] desenvolveu uma ferramenta de apoio ao piloto para auto-separação

baseada em estado para aeronaves em voo de cruzeiro. O modelo considera as restrições de manobras táticas em um espaço de ação do vetor de direção de velocidade. Para evitar o cálculo de manobras fora dos limites de desempenho da própria aeronave e do tráfego circundante foi remodelada a típica geometria de conflito usada para apresentar restrições de manobra tática da própria aeronave. A técnica *break-circle* foi usada para determinar se as manobras da aeronave farão com que ambas as aeronaves passem uma pela outra com a separação desejada após a realização das manobras. Também foi discutido como essas restrições de manobra podem ser apresentadas no sistema de gerenciamento de voo para que os pilotos possam estar cientes do efeito das mudanças de controle do modo de aeronave nas restrições. O trabalho resultou em um conceito de exibição de intenção que ajuda os pilotos a lidar efetivamente com situações de conflito baseadas em estado e baseadas em intenção em diferentes modos de controle de aeronaves (figura 4.4). Entretanto não são apresentados resultados quantitativos que permitam uma avaliação objetiva dos benefícios oferecidos pelo modelo proposto.

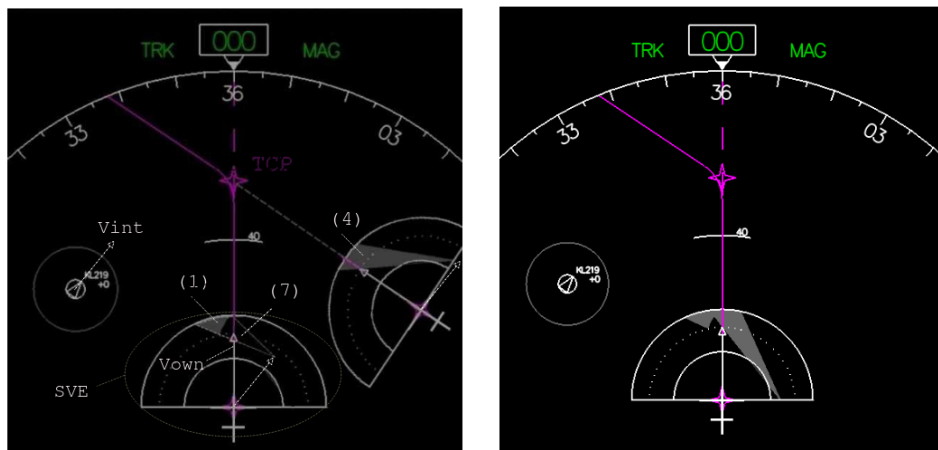


Figura 4.4: Apresentação do método *break-circle* para a tripulação[63].

Uma representação denominada obstáculo velocidade (VO) foi utilizada por Velasco *et al.* [64] como uma abstração matemática cujo objetivo é garantir a separação para a navegação autônoma de veículos. Um VO é definido como o conjunto de todos os vetores de velocidade de uma aeronave em movimento que resultará em uma colisão com um obstáculo em movimento em algum momento, assumindo que o obstáculo em movimento mantenha um vetor de velocidade constante. O trabalho busca evitar não somente colisões, mas as perdas de separação, onde uma área circular é assumida em torno de cada aeronave, que deve permanecer livre de outros objetos. Essa representação apresenta desempenho ainda melhor quando a intenção de trajetória é compartilhada entre as aeronaves envolvidas no conflito.

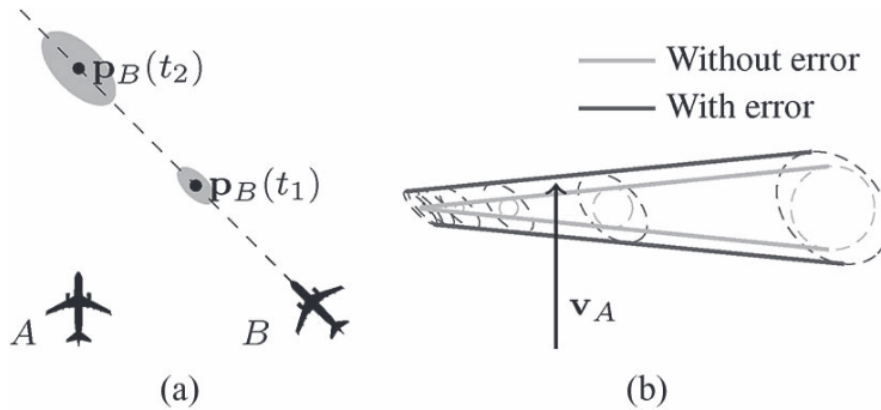


Figura 4.5: Ilustração do método obstáculo velocidade (VO) [64].

O trabalho propôs uma fórmula analítica capaz de incorporar conceitos variáveis no tempo, como erros de previsão de trajetória. Estudos para situações específicas foram explorados e as possibilidades de estender o método à terceira dimensão foram discutidas. A análise mostrou que o cálculo do VO com informações de intenção pode sugerir, de forma satisfatória, manobras evasivas a uma curta distância. Este autor também não apresentou resultados quantitativos, inviabilizando uma análise mais detalhada dos resultados obtidos.

4.2 Métodos de CDR Utilizando Aprendizado de Reforço Profundo

Ferramentas de aprendizado de reforço profundo tem sido utilizadas para diversas pesquisas para solucionar problemas de CDR entre outros relacionados ao Gerenciamento de Tráfego Aéreo [65, 66, 67]. Entre elas, destaca-se o TensorFlow, um sistema de aprendizado de máquina que opera em larga escala e em ambientes heterogêneos. Trata-se de uma biblioteca de software gratuita e de código aberto que pode ser aplicada em tarefas de aprendizado de máquina e inteligência artificial. Caracteriza-se pela possibilidade de poder ser utilizado em uma variedade de tarefas, mas tem um foco particular no treinamento e inferência de redes neurais profundas. Ele mapeia os nós de um gráfico de fluxo de dados em várias máquinas em um cluster, mas também dentro de uma máquina em vários dispositivos computacionais, incluindo CPUs multicore, GPUs de uso geral e ASICs personalizados conhecidos como unidades de processamento tensor [68].

Brittain e Wei fizeram contribuições significativas para o campo de pesquisa de CDR utilizando aprendizado de reforço profundo. Em [69], uma estrutura hierárquica de apren-

dizado por reforço profundo é usada para sequenciar e separar aeronaves em rota, ajustando sua velocidade em um ambiente 2D. A entrada para o DRL é composta por pixels brutos da tela e uma matriz de dados de todos os voos. Uma Double DQN é usada. Seu sistema alcançou a solução ideal após cerca de 7.000 episódios de treinamento em um cenário que incluía duas aeronaves com duas opções de rota para cada uma. A figura 4.6 mostra a estrutura de rotas, cujas imagens foram fornecidas como entrada para a rede Double DQN, representada como uma camada oculta na imagem. Este trabalho utiliza rotas fixas, ou seja, o sistema de gerenciamento de rotas da aeronave tem a opção de escolha entre rotas pré-definidas. O modelo proposto pode ser aplicado com mais facilidade ao cenário atual de ATM porém não é possível uma aplicação direta para o cenário do UAM com rotas livres.

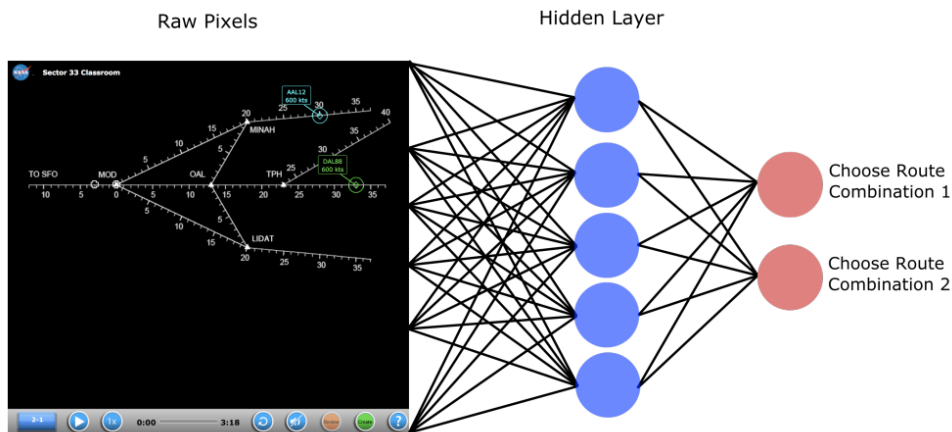


Figura 4.6: Ilustração da arquitetura aplicada no trabalho de Brittain e Wei [69].

Em [70], uma estrutura de DRL multiagente foi proposta para detectar e resolver conflitos em 2D composto por um setor em rota com múltiplas interseções e pontos de fusão. O controle das aeronaves é feito com ajustes de velocidade. A estrutura usa um modelo ator-crítico para aprendizagem centralizada, mas a execução é descentralizada. O modelo resolveu todos os conflitos em 99,97% dos testes após 20.000 episódios em um estudo de caso contendo 30 aeronaves com três rotas e duas interseções, representado na figura 4.7. Em [71], resultados promissores foram obtidos usando uma estrutura chamada Deep Distributed Multi-Agent Variable que utiliza um algoritmo ator-crítico que incorpora uma Long Short-Term Memory (LSTM). Todos os resultados deste grupo de pesquisa foram validados no simulador BlueSky [72]. A resolução de conflitos por meio de ajuste de velocidade aplicada nos dois trabalhos descritos neste parágrafo é uma solução importante, mas deixa de lado outras manobras que podem ser mais efetivas para a resolução do conflito, incluindo a mudança de direção das aeronaves envolvidas.



Figura 4.7: Ilustração das rotas aplicadas por Brittain e Wei [70].

Pham *et al.* [73] apresentou um agente capaz de resolver conflitos e lidar com incertezas nas manobras de resolução de conflitos sem a necessidade de conhecimento prévio sobre um conjunto de regras mapeando de cenários de conflito para ações esperadas. Este sistema de agente único opera em espaço aéreo que não possui uma estrutura pré-definida. A tarefa de resolução de conflitos é formulada como um problema de tomada de decisão no espaço de ação grande e complexo, aplicável ao empregar o algoritmo de aprendizado de reforço. O trabalho inclui o desenvolvimento de um ambiente de aprendizado, representação do estado de cenário, função de recompensa e algoritmo de aprendizado. As manobras são limitadas a mudanças de direção e, portanto, funcionam em um ambiente bidimensional (figura 4.8). Seu modelo foi capaz de resolver conflitos com uma taxa de sucesso de mais de 81%. O objetivo deste trabalho é uma ferramenta de apoio à decisão do ATCo, cuja aplicação para a resolução de conflitos entre aeronaves eVTOL num cenário em que poderá não haver o gerenciamento dos tráfegos por um Controlador requer adaptação.

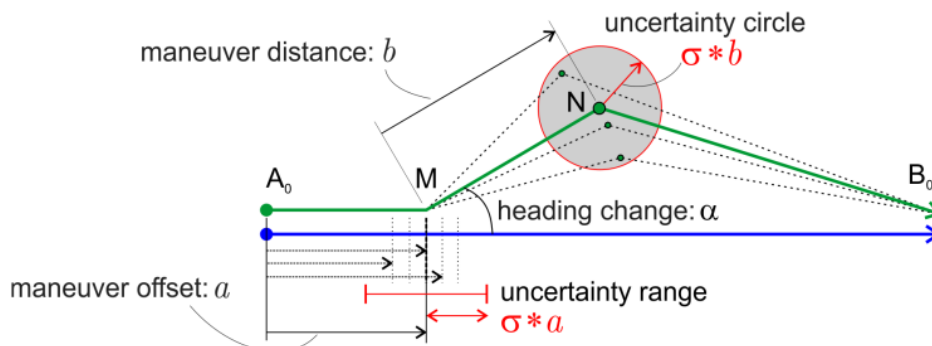


Figura 4.8: Pham *et al.* [73] utilizou um sistema capaz de lidar com incertezas nas manobras para evitar conflitos.

O modelo desenvolvido por Li *et al.* [74] usa uma abordagem em duas etapas. Na primeira etapa, eles usaram programação dinâmica para calcular as políticas de resolução de conflitos em pares. Em seguida, a decomposição de utilidade é aplicada para aproximar os procedimentos ideais para resolução de conflitos de múltiplas ameaças. A segunda etapa consiste em uma correção para operações em espaço aéreo denso aprendida por uma rede DQN. Os resultados foram avaliados em termos de segurança e eficiência. O modelo de melhor desempenho em segurança manteve a média de colisões abaixo de 6 por hora de voo ($\times 10^{-3}$) em espaços aéreos contendo até 40 voos/ km^2 . Para avaliação da eficiência, eles mediram o comprimento normalizado da rota, definido como a razão entre o comprimento do trajeto voado e a distância do início ao destino. A média do comprimento da rota normalizada variou entre um adicional de 10% a 60% a depender da densidade de aeronaves utilizada na simulação. O modelo de geração de trajetórias aplicado no teste de *stress* faz com que todas as aeronaves voem em direção ao mesmo ponto, para maximizar a possibilidade de conflito. Esta estratégia tem pouca relação com o comportamento de aeronaves no mundo real.

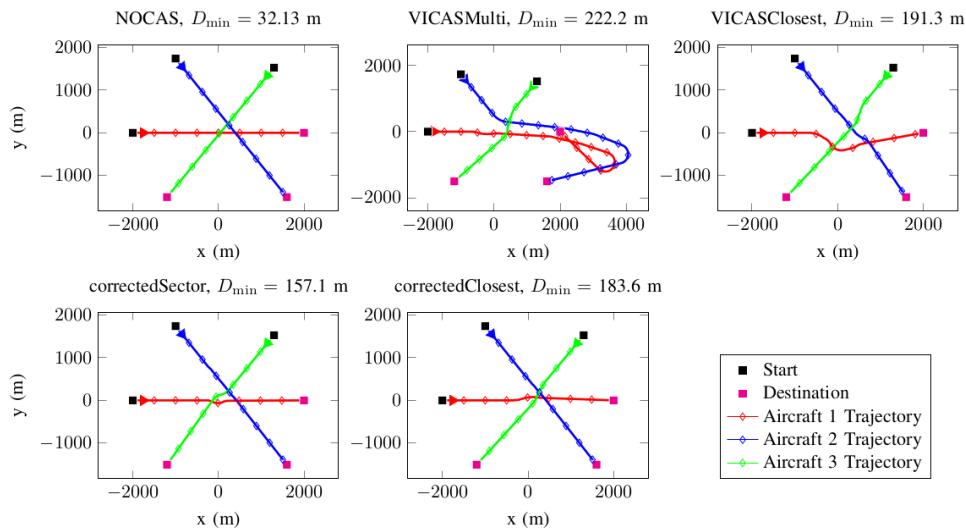


Figura 4.9: Resultados da aplicação dos modelos propostos por Li et al. [74].

Ribeiro *et al.* [75] emprega um método de aprendizado por reforço (RL) responsável por decidir os parâmetros que um método geométrico de resolução de conflitos (CR) poderá utilizar para gerar a manobra para solucionar cada situação de conflito. Os resultados mostram que esta abordagem híbrida, combinando os pontos fortes dos métodos geométricos CR e RL, reduz o número total de perdas de separação mínima. Além disso, a grande variedade de diferentes soluções ótimas encontradas pelo método RL mostra que as regras do método CR geométrico devem ser expandidas, atendendo a diferentes geometrias de conflito. O modelo proposto é denominado Soft Actor–Critic (SAC), que treina dentro do ambiente do espaço aéreo para encontrar os valores ótimos para o cálculo de

uma manobra de resolução de conflitos. Experimentos são conduzidos com a ferramenta de simulação ATC multiagente de código aberto BlueSky [72]. Entretanto, não é possível concluir se a proposta será efetiva em cenários reais considerando as incertezas que podem surgir relacionadas à meteorologia e problemas na comunicação.

Uma rede Dueling-DQN foi usada por Nunes *et al.* [76] para criar um modelo focado na explicabilidade e conformidade da IA para aumentar a aceitação da solução proposta aos ATCOs. Embora os objetivos do trabalho deles sejam diferentes dos nossos, este é o único modelo que usa o mesmo tipo de rede que este trabalho para a resolução de conflitos.

4.3 Discussão

Os artigos citados acima resumem as pesquisas que já foram feitas no campo que este trabalho está inserido. As cinco primeiras referências da tabela 7.2 consideram aeronaves evoluindo em um espaço aéreo de rotas livres (FRA) mas não utilizam métodos de DRL. Adicionalmente, apenas as referências [60] e [61] apresentaram resultados quantitativos que podem ser comparados com outros trabalhos. Entre eles, os resultados de Hoekstra *et al.* foram mais significativos. Entretanto, como as resoluções deste trabalho são baseadas apenas na geometria do conflito, elas podem se opor à direção de voo desejada, levando a um aumento na distância percorrida e consequente aumento no consumo de combustível. Assim sendo, nenhum dos trabalhos citados neste parágrafo levou em consideração a eficiência da rota empregada na resolução dos conflitos.

Tabela 4.1: Resumo dos Trabalhos Relacionados

Referência	Autor(es)	FRA	DRL	Resultados Quantitativos	Eficiência da Rota
[60]	Hoekstra <i>et al.</i>	✓	X	✓	X
[61]	Balasooryan	✓	X	✓	X
[62]	Paielli	✓	X	✓	X
[63]	Van Dam <i>et al.</i>	✓	X	X	X
[64]	Velasco <i>et al.</i>	✓	X	X	X
[69]	Brittain e Wei	X	✓	✓	N/A
[70]	Brittain e Wei	X	✓	✓	N/A
[71]	Brittain	X	✓	✓	N/A
[73]	Pham <i>et al.</i>	X	✓	✓	N/A
[74]	Li <i>et al.</i>	✓	✓	✓	✓
[75]	Ribeiro <i>et al.</i>	✓	✓	✓	X
Presente Trabalho		✓	✓	✓	✓

As demais referências apresentadas na tabela 7.2 utilizam DRL para solução de conflitos. Apesar de terem apresentado bons resultados, os trabalhos [69], [70] e [71] consideram um espaço aéreo de rotas pré-definidas, no qual não é possível avaliar a eficiência do método em relação a deslocamento para solução do conflito. A referência [73] utiliza uma estrutura do espaço aéreo que permite à aeronave realizar desvios da rota para solucionar conflitos, porém não foi feita nenhuma análise da eficiência dos desvios realizados.

O artigo apresentado na referência [74] apresenta bons resultados e tem bastante interseção com o presente trabalho. Porém não foi realizada uma comparação entre diferentes modelos de DRL, o que poderia ter apresentado resultados ainda melhores. Por fim, a referência [75] apresentou excelentes resultados e análises, mas não considerou a eficiência dos desvios realizados para a solução de conflitos.

Portanto, a análise realizada mostra que ainda existem algumas perguntas que não foram respondidas quando se trata de aplicação de métodos de DRL em um FRA para gerenciamento de trajetórias em caso de conflitos. Este trabalho busca responder algumas das questões que ainda não foram abordadas na literatura.

Capítulo 5

Metodologia e Modelagem

Uma vez que os conceitos básicos foram apresentados nos capítulos anteriores, neste capítulo o problema sob análise será definido de forma mais precisa. Em seguida será enunciada a proposta de solução que será aplicada neste trabalho, seguida da metodologia empregada. Por fim, será abordada a modelagem que foi utilizada.

5.1 Definição do Problema

O cenário para qual este trabalho está sendo desenvolvido é um espaço aéreo no qual existirão diversas aeronaves eVTOL em voo simultâneo, pilotadas ou autônomas. Neste cenário, serão utilizados sistemas redundantes para detecção e resolução de conflitos. O primeiro destes sistemas, cuja implementação foge ao escopo deste texto, entrará em ação no momento da análise do plano de voo, ou seja, buscando realizar a resolução estratégica dos conflitos. Nesta fase, caso um conflito seja detectado, alterações serão realizadas no plano de voo de modo a obter uma trajetória livre de conflitos, seja por alteração nas rotas ou por modificação dos horários previstos para passagem nos pontos de cruzamento entre as aeronaves.

Cada aeronave poderá seguir voo livre conforme descrito em 3.3.3, em trajetória direta até o seu destino, executando desvios somente se houver algum conflito. A hipótese é de que não haverá um controle centralizado, portanto a cada aeronave deverá ser capaz de detectar eventuais conflitos durante o voo e realizar manobras para evitá-lo.

Mesmo após a desconflitação estratégica, é possível que fatores imprevistos afetem a rota das aeronaves e também a execução do voo de acordo com os horários autorizados. Assim, é necessário que um sistema de resolução de conflitos tático também esteja em funcionamento para definir manobras que coloquem as aeronaves em novas trajetórias livres de conflitos, bem como evitar que novos conflitos sejam criados no processo.

Portanto, o problema para o qual será proposta uma solução neste trabalho é o desenvolvimento de um sistema capaz de determinar desvios táticos de trajetórias entre aeronaves eVTOL em caso de conflitos. O suposto problema se passa no ambiente UAM de modo que o objetivo da aplicação do sistema é permitir que as aeronaves possam executar voo livre durante a maior parte de seus deslocamentos.

O cenário descrito pode ser visto como um problema computacional, no qual existe uma quantidade infinita de configurações para os conflitos envolvendo duas ou mais aeronaves e a necessidade de definição de qual seria a melhor rota a ser tomada em cada um dos possíveis conflitos.

5.2 Proposta de Solução

O sistema proposto, denominado ISUAM, visa utilizar modelos de aprendizado profundo, cujas redes, após um período de treinamento, seriam capazes de oferecer como respostas os desvios necessários para a resolução dos conflitos. Tal solução se dará na fase tática de resolução de conflitos, conforme explicitado na Figura 3.6. O esquema geral da proposta é apresentado na Figura 5.1 e será descrito abaixo.

A fase de treinamento será composta de uma série de episódios. No início de cada episódio, um módulo denominado Gerador de Ambiente aciona um submódulo que definirá a posição dos vertiportos que serão empregados e, em seguida, essas informações são utilizadas para gerar todos os voos que serão simulados. O ambiente gerado será utilizado em cada um dos passos do episódio conforme descrição a seguir.

Para cada passo do episódio, um módulo de detecção de conflitos é aplicado. Caso seja detectado algum conflito, as informações que caracterizam a posição e os ângulos entre as aeronaves envolvidas, também chamada de geometria do conflito, são repassadas para o ISUAM que está sendo utilizado e também para os gerenciadores das estratégias alternativas.

O ISUAM recebe a geometria do conflito e responde com uma ação sob a forma de uma direção a ser tomada pela aeronave a partir da sua direção atual. O sistema de gerenciamento da simulação executa a mudança de direção e o resultado é fornecido para um sistema que avaliará a qualidade da resolução do conflito e oferecer uma recompensa conforme política pré-definida.

Após a realização de todos os episódios de treinamento, o sistema convergirá para uma configuração estável que responderá da forma mais adequada possível na fase de aplicação. Nessa fase, as informações do conflito real serão fornecidas para o ISUAM pré-treinado que, por sua vez, responderá com a direção mais adequada a ser tomada para a solução do conflito.

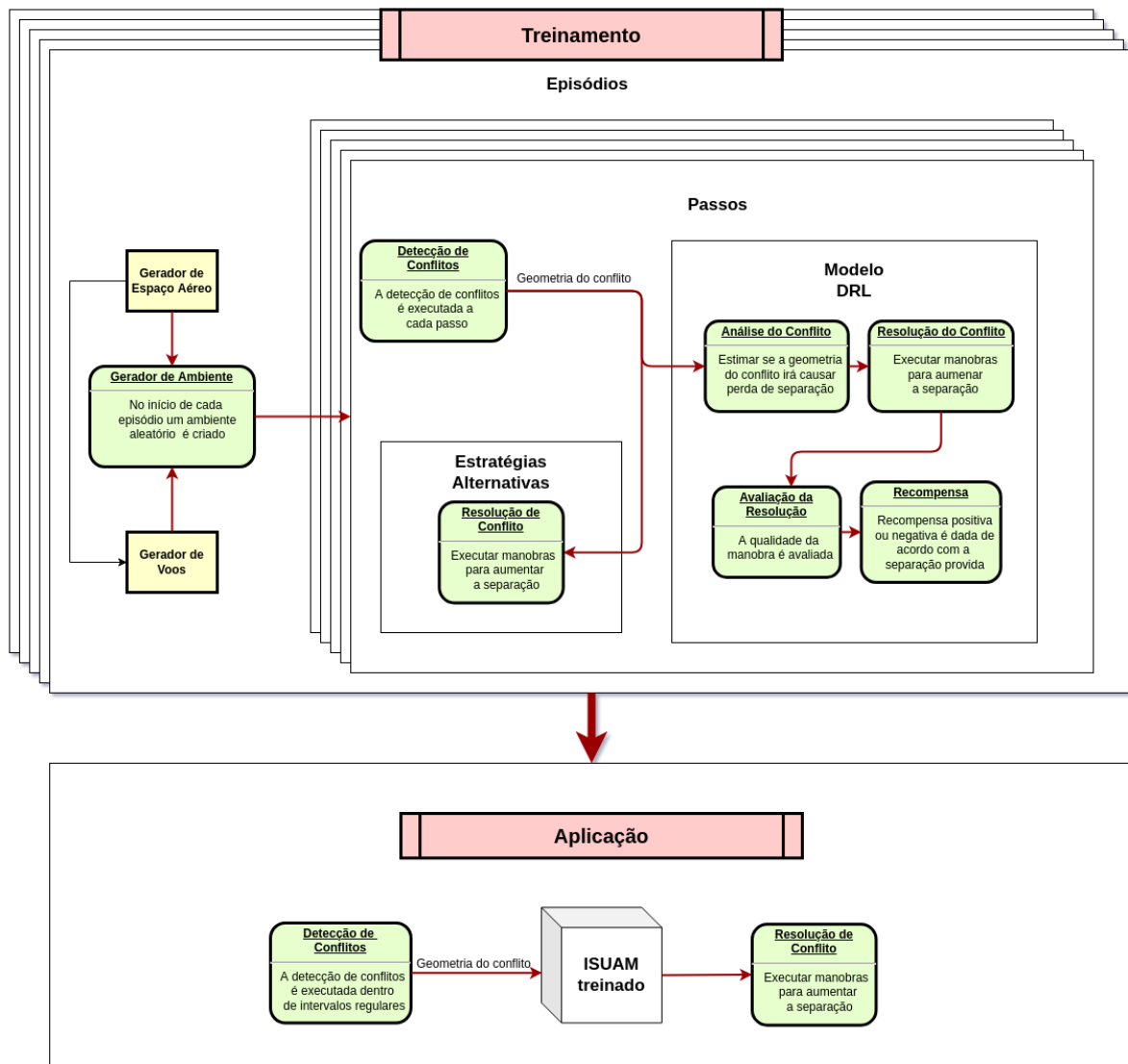


Figura 5.1: Representação da fase de treinamento e de aplicação do ISUAM.

5.3 Metodologia

Como foi visto na seção 2.3.3, os modelos de aprendizado de reforço são capazes de aprender quais são as ações mais adequadas a tomar em cada situação, em outras palavras, como mapear situações para ações. Durante o período de treinamento, o sistema é capaz de descobrir qual as ações rendem a maior recompensa ao serem tomadas. Além disso, as ações podem afetar não apenas a recompensa imediata, mas também a próxima situação e, através disso, todas as recompensas subsequentes. As características mencionadas acima, tentativa e erro busca e recompensa atrasada, tornam este tipo de modelo vantajoso para a presente aplicação [77].

5.4 Modelagem

A solução de conflitos será modelada como um Processo de Decisão de Markov, conforme notação da Seção 2.1.

- O espaço de estados \mathcal{S} inclui todas as possíveis configurações de posições e das aeronaves em evolução num determinado espaço aéreo;
- O espaço de ação \mathcal{A} que consiste em todas as ações que o agente pode realizar, neste caso são as decisões de realizar curvas;
- Função de transição $\mathcal{T}(s_{t+1}|s_t, a_t)$ que descreve a probabilidade de chegar ao estado s_{t+1} , dado o estado atual s_t e a ação a_t .
- A função de recompensa $\mathcal{R}(s_t, a_t, s_{t+1})$ que decide a recompensa imediata recebida após a transição do estado s_t para o estado s_{t+1} , devido à ação a_t . Um dos objetivos deste trabalho é buscar valores adequados para a função de recompensa que façam com que o agente apresente o comportamento desejado.
- Um fator de desconto $\gamma \in [0, 1]$ que decide a preferência entre a recompensa imediata versus recompensas futuras. No problema em questão o valor também será definido por meio de experimentos.

O sistema será treinado para encontrar uma política ótima π^* que, caso seja seguida a partir de qualquer estado inicial, maximiza a recompensa cumulativa esperada para todos os passos futuros:

$$\pi^* = \operatorname{argmax}_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} R(s_t, a_t, s_{t+1}) \mid \pi\right] \quad (5.1)$$

A função- Q , caso dos modelos de DRL, não é apresentada de forma explícita, mas pode ser obtida de forma indireta após o treinamento do modelo, por meio das configurações da rede.

A função de valor ótimo $V^*(s)$ denotará a recompensa total máxima esperada ao iniciar do estado s , que pode ser expresso como o máximo de $Q^*(s, a)$ sobre todas as ações possíveis:

$$V^*(s) = \max_a Q^*(s, a) \quad \forall s \in \mathcal{S} \quad (5.2)$$

5.4.1 Conflitos

Abaixo serão modeladas as funções que serão aplicadas para indicar a existência ou a resolução de conflitos.

Seja $d_h^{A_i A_j}(t)$ a distância horizontal entre as aeronaves A_i e A_j no instante t , e seja $d_v^{A_i A_j}(t)$ a distância vertical entre as aeronaves A_i e A_j no instante t onde S_h e S_v são, respectivamente, a separação horizontal e vertical mínima exigida.

Então, é possível afirmar que um conflito c entre as aeronaves A_i e A_j no instante t é definido por:

$$c^{A_i A_j}(t) \leftrightarrow (d_h^{A_i A_j}(t) < S_h) \wedge (d_v^{A_i A_j}(t) < S_v) \quad (5.3)$$

Seja $I_c(t)$ a função que indica se existe um conflito entre duas aeronaves no instante t definida por:

$$I_c(t) = \begin{cases} 1 & \text{se } c^{A_i A_j}(t), \\ 0 & \text{se } \neg c^{A_i A_j}(t). \end{cases} \quad (5.4)$$

Penalidade por Conflitos

A existência de conflitos em um determinado passo fará com que a recompensa oferecida seja negativa conforme descrito abaixo.

Seja $r_c^-(t)$ a recompensa negativa recebida no passo t caso existam conflitos naquele passo e T a duração total do episódio. Assim, a soma das recompensas negativas recebidas pelo agente caso exista um conflito entre duas aeronaves em todos os passos do episódio pode ser definida por:

$$R_c^- = \sum_{t=0}^{T-1} r_c^-(t) I_c(t). \quad (5.5)$$

Resolução de Conflitos

A capacidade do agente de resolver conflitos levará a uma recompensa positiva de acordo com a formalização a seguir.

Seja $I_r(t)$ a função que indica se um conflito que existia no instante $t - 1$ foi resolvido no instante t definida por:

$$I_r(t) = \begin{cases} 1 & \text{se } I_c(t-1) \wedge \neg I_c(t), \\ 0 & \text{se caso contrário.} \end{cases} \quad (5.6)$$

Seja $r_r^+(t)$ a recompensa positiva recebida no passo t caso de um conflito existente no passo anterior tenha sido solucionado e T a duração total do episódio. Assim, a soma das

recompensas positivas recebidas pelo agente por soluções de conflitos pode ser definida por:

$$R_r^+ = \sum_{t=1}^{T-1} r_r^+(t)I_r(t). \quad (5.7)$$

5.4.2 Eficiência dos Desvios

As operações relativas a direções realizadas nesta seção consideram as medidas em radianos, portanto somas que ultrapassem o valor de 2π ou subtrações que tenham valor inferior a 0 são possíveis. Nos casos em que for necessário realizar a conversão de x radianos para graus, a operação utilizada é $deg(x)$.

Considerando que a direção a na qual a aeronave esteja voando em um determinado instante, a mudança de direção m determinada pelo agente DRL e d a direção para o destino da aeronave, então a recompensa negativa a ser recebida no instante t é dada por:

$$r_d^-(a, m, d, t) = -\lfloor deg(|a + m - d|)/10 \rfloor \quad (5.8)$$

.

Ou seja, cada 10 graus de diferença entre a direção assumida pela aeronave após um desvio e a direção do destino correspondem a uma unidade de recompensa negativa.

Entretanto, a aeronave não executará desvios em todos os passos de um episódio. Assim, a função $I_d(t)$ que será utilizada para indicar a existência de um desvio no instante t é definida por:

$$I_d(t) = \begin{cases} 1 & \text{se } m \neq 0, \\ 0 & \text{se } m = 0. \end{cases} \quad (5.9)$$

Portanto, a soma das recompensas negativas recebidas pelo agente em função dos desvios realizados pode ser definida por:

$$R_d^- = \sum_{t=0}^{T-1} r_d^-(a, m, d, t)I_d(t). \quad (5.10)$$

5.4.3 Função de Maximização

A partir das somas descritas nas seções anteriores, é possível definir o objetivo da política ótima utilizando a notação da Equação 5.1 na qual o objetivo do agente de DRL a cada episódio é a maximização das somas das funções da seguinte forma:

$$\pi^* = \operatorname{argmax}_{\pi} \mathbb{E} \left[\sum_{t=0}^{T-1} -R_c^- + R_r^+ - R_d^- \mid \pi \right] \quad (5.11)$$

Cabe observar que as somas de recompensas R_c^- e R_d^- foram aplicadas na Equação 5.11 com sinal negativo para que a maximização tenha sempre sinal positivo.

Capítulo 6

Implementação

Este capítulo descreve o simulador que foi desenvolvido para testar a capacidade de resolução de conflitos por modelos de aprendizado de reforço profundo no ambiente UAM. Diferente de outros simuladores já existentes, optou-se por criar uma implementação customizada visando futuras aplicações no DECEA. A solução foi desenvolvida em Python e permite configurar o espaço aéreo simulado, a quantidade de vertiports e aeronaves, e serão realizados testes com diferentes parâmetros ajustáveis. O simulador estabelece zonas de proteção ao redor de cada aeronave, com uma zona de monitoramento externa, uma zona de conflito intermediária e uma zona crítica interna. O objetivo é evitar conflitos e realizar desvios apenas quando necessário. Estratégias de comparação foram estabelecidas para um melhor parâmetro de comparação com o ISUAM. O capítulo encerra com a descrição da função de recompensa utilizada no treinamento.

6.1 Descrição da Implementação

Um simulador customizado foi desenvolvido com o objetivo de testar a capacidade de resolução de conflitos por parte dos modelos de aprendizado de reforço profundo no ambiente UAM. Entre os simuladores já existentes, destaca-se o Bluesky [72] por ser um simulador de código aberto que já foi utilizado em diversos trabalhos na literatura. Porém a opção de criação de uma implementação customizada se deu devido ao fato de que o produto poderia ser disponibilizado para outras aplicações do DECEA em pesquisas futuras.

A linguagem escolhida para o desenvolvimento do simulador foi Python, devido à grande quantidade de bibliotecas para aprendizado de máquina já disponíveis. Neste caso foi utilizada a versão 0.6.5 do Tensorforce, uma implementação baseada no TensorFlow [68]. O código utilizado neste trabalho foi desenvolvido a partir do trabalho de Berthelot [78] e está disponível no GitHub [79].

O simulador pode ser configurado de acordo com a demanda sob vários aspectos. Entre eles, estão a dimensão do espaço aéreo simulado, a quantidade de vertiports e a quantidade de aeronaves envolvidas na simulação. Serão realizados testes com diversos valores para os parâmetros que podem ser ajustados no sistema.

Foram estabelecidas zonas de proteção ao redor de cada aeronave conforme a figura 6.1. A região externa é uma zona de monitoramento, cujo objetivo é apenas determinar se existe a possibilidade de um intruso ingressar na zona intermediária, que é a zona de conflito. As recompensas no treinamento serão estabelecidas de modo a evitar que qualquer intruso ingresse nessa zona. Entretanto, o ingresso na zona de conflito não implica, por si só, em uma colisão. Para isso foi estabelecida uma zona interna, denominada zona crítica. O ingresso de um intruso na zona crítica significa colisão ou uma separação abaixo dos valores aceitáveis.

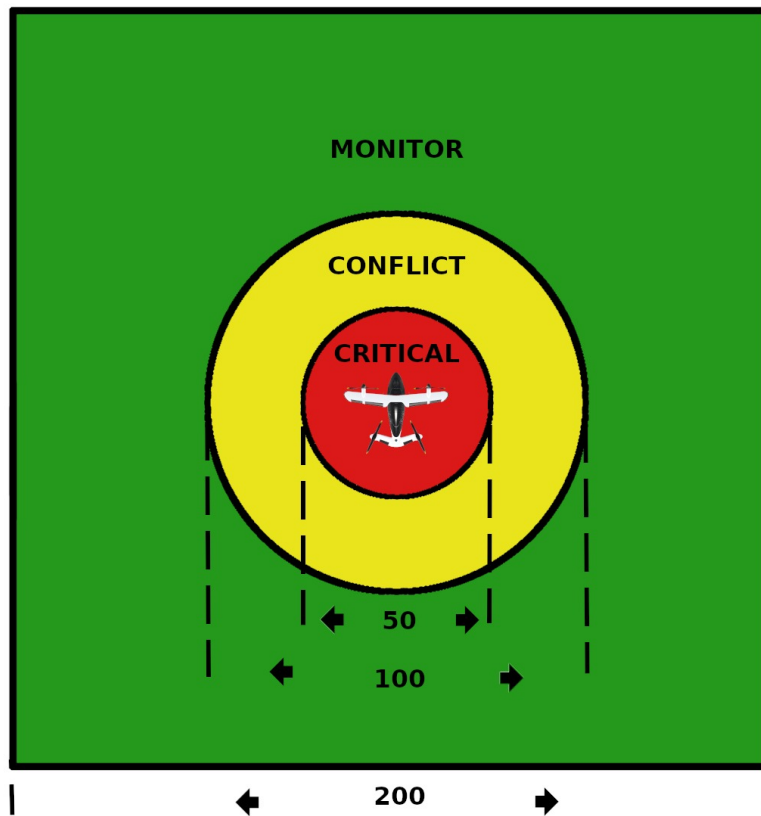


Figura 6.1: Zonas de proteção ao redor de cada aeronave.

Com o objetivo de aumentar a probabilidade de conflitos e testar o potencial do sistema, todas as aeronaves envolvidas em uma simulação serão consideradas como se estivessem navegando na mesma altitude. Ou seja, toda a simulação ocorre apenas em duas

dimensões. Em caso de conflito, o sistema determinará a realização de curvas à direita ou à esquerda, com o objetivo de realizar desvios da rota original que resolvam o conflito.

6.2 Estratégias para Comparação

Foram implementadas outras estratégias de resolução de conflitos para obtenção de parâmetros de comparação com o modelo ISUAM. Abaixo serão descritas as estratégias que objetivam evitar o conflito a todo custo, realizar manobras aleatórias e chegar ao destino sem desvios, ignorando conflitos.

6.2.1 Avoid

A estratégia denominada *Avoid* busca evitar um conflito a todo custo. Para isso, sempre que há um intruso na zona de monitoramento, esta estratégia determina a realização de manobras para que a aeronave voe em direção oposta. A figura 6.2 mostra a sugestão desta estratégia para solucionar um conflito com um intruso localizado na posição B. Uma vez que o intruso está posicionado a 50 graus à esquerda da aeronave, a estratégia determina a realização de uma curva de 130 graus à direita. Fazendo com que a aeronave voe na direção oposta à localização do intruso.

Na comparação com a estratégia DRL, o objetivo é que a estratégia seja um marcador do melhor valor que seria possível obter em termos de resolução de conflitos. Em relação à eficiência da estratégia, é esperado que o *Avoid* tenha o pior resultado e que o DRL tenha resultados melhores.

6.2.2 Random

A estratégia denominada *Random* retorna determinações de curvas de modo aleatório. Estas curvas têm uma faixa de possibilidades entre -90 e +90, com variação de 10 unidades entre si. Onde -90 significa uma curva de 90 graus à esquerda e +90 significa uma curva de 90 graus à direita. Os modelos de DRL terão um comportamento esperado muito próximo desta estratégia durante o início do treinamento. Porém espera-se que com o passar dos episódios os valores passem a se afastar dessa referência.

6.2.3 Free

Por fim, a estratégia denominada *Free* executa o voo livre, seguindo em direção ao destino e ignorando possíveis conflitos na trajetória. Esta estratégia possivelmente será a mais eficiente em termos de deslocamento, visto que não há desvios sob nenhuma hipótese,

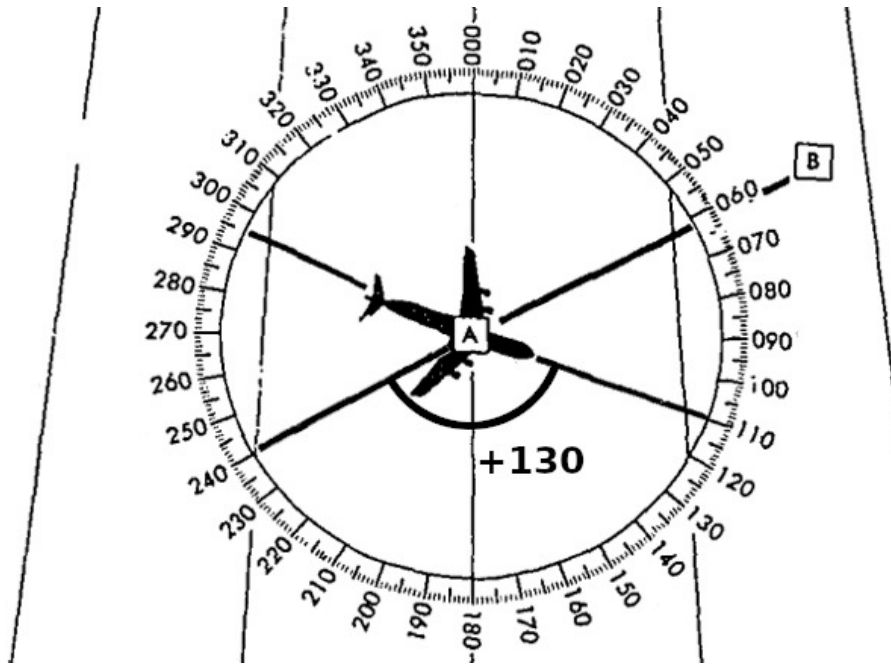


Figura 6.2: Curva sugerida pela estratégia avoid para evitar um conflito.

porém ao custo de eventuais perdas de separação. Entretanto, o resultado esperado é de que esta estratégia tenha os piores resultados na resolução de conflitos.

6.3 Recompensa

Para garantir que o sistema seja capaz de definir quais seriam as melhores decisões em cada situação, é necessário fornecer dados suficientes e avaliar as ações do agente de forma adequada. Nesta implementação, as informações fornecidas para o agente, ou seja, o estado do ambiente são compostas por uma matriz que contém, para cada intruso, a distância, a posição relativa e a proa relativa. Também é fornecida a direção do destino da aeronave.

A recompensa utilizada é composta por três contribuições: a capacidade do agente de evitar conflitos, a capacidade de resolver os conflitos existentes e a capacidade de realizar desvios sem um afastamento desnecessário da rota ideal. Caso um conflito já esteja em andamento, o agente recebe uma recompensa positiva, de valor +1 se for capaz de solucioná-lo. Se o conflito não for resolvido e o intruso ingressar na zona de conflito, a recompensa será negativa, com valor -5 , ao passo que se o intruso ingressar na zona crítica, a recompensa será -10 . A situação desejada é a inexistência de conflitos, por isso caso essa situação seja detectada, a recompensa terá o valor de +10.

Em relação à eficiência do desvio, é efetuado um cálculo que compara a direção na qual a aeronave está voando em função do desvio realizado com a direção do destino. O valor em graus é arredondado para o múltiplo de 10 mais próximo. Por exemplo, 38 graus é arredondado para 40 graus e esse valor, por sua vez, é dividido por dez para obter o valor 4. Nesse caso, o valor positivo de 38 graus indica que o destino estaria 38 graus à direita da direção atual da aeronave. Por isso a recompensa será de -4 conforme o cálculo descrito.

Dessa forma os valores para esse componente oscilarão entre -18 e $+18$ sendo que não haverá recompensa negativa se a aeronave estiver voando na direção ideal do seu destino.

Capítulo 7

Experimentos

Este capítulo relatará a configuração padrão e os testes realizados a partir da implementação descrita anteriormente. Foram realizados testes para determinar a duração necessária do treinamento, para comparar os diferentes modelos de DRL, os efeitos da densidade de tráfegos no desempenho e também a eficiência em relação ao aumento do deslocamento para a solução de conflitos.

7.1 Configuração

A simulação foi realizada em um servidor com processador Intel(R) Core(TM) i5-9400F CPU @ 2.90GHz com 6 núcleos e equipado com 32 GB de memória RAM. Os modelos de DRL disponibilizados pelo Tensorforce foram aplicados com as configurações padrão que já são pré-definidas na biblioteca.

Foi utilizado um espaço aéreo de dimensões 1000 m x 1000 m no qual, a cada episódio, são colocados 30 vertiports em posições aleatórias. O número de aeronaves utilizado na maioria dos experimentos foi de 7 por episódio, exceto pelo teste de densidade, onde esse número também foi alterado para avaliação dos resultados. O número de sete eVTOL a cada episódio foi definido em testes preliminares por oferecer uma quantidade mediana de conflitos, nos quais seria possível avaliar o desempenho do modelo sem congestionamento excessivo.

No início dos episódios, cada eVTOL é designado para decolar de um dos vertiports de forma aleatória. Da mesma forma, é selecionado como destino um dos demais vertiports. Essa forma de colocação das posições de origem e destino implica que o modelo será treinado para lidar praticamente com todas as geometrias de conflito possíveis desde que o número de episódios seja grande o suficiente.

O deslocamento entre o vertiporto de origem e destino é dividido em 30 passos a cada episódio. Isso significa que aeronaves com deslocamentos mais curtos devido à pro-

ximidade do par origem/destino voam com velocidades menores. Este valor também foi definido em testes preliminares por permitir a realização de desvios quando necessário sem tornar o episódio demasiadamente longo, o que afetaria o tempo de treinamento total do modelo.

A cada episódio o modelo de DRL que está sob análise é executado em paralelo com as estratégias alternativas descritas na subseção 6.2. Isso garante que a comparação seja feita em condições de igualdade, visto que todas as estratégias estão enfrentando exatamente o mesmo cenário de conflitos.

7.2 Ilustração de um Episódio

Sempre que um conflito é detectado, a geometria do conflito é repassada como entrada para o modelo DRL para que o mesmo decida qual vai ser a ação tomada pela aeronave. Logo após a resolução do conflito, a aeronave retoma o voo direto para alcançar o seu destino. Para melhor visualização, a Figura 7.1 ilustra todos os passos dados em um episódio, com seus respectivos rastros, no qual ocorrem dois conflitos, destacados pelo círculo verde e vermelho na região central da imagem.

Tabela 7.1: Aeronaves presentes no episódio de ilustração

Cor do rastro	Ponto de origem	Ponto de destino
Vermelho	A	B
Roxo	C	D
Azul	E	F
Laranja	G	H
Verde	I	J

Nesse episódio foi utilizado um modelo de DRL já treinado de acordo com os parâmetros que serão descritos neste capítulo. A figura mostra o rastro da trajetória de cinco aeronaves cujas cores, origem e destino são apresentadas na Tabela 7.1. Para facilitar a compreensão, as aeronaves serão denominadas de acordo com a cor de seu rastro na imagem. Os dois conflitos que aparecem na imagem serão descritos em maiores detalhes abaixo.

A Figura 7.2a mostra o momento do início do conflito 1, no qual as duas aeronaves estão se aproximando e o sistema detecta que a aeronave roxa ingressará na zona de monitoramento da aeronave azul. Essa detecção é indicada pela cor vermelha na posição onde está a aeronave roxa neste momento. Na Figura 7.2b é apresentado o momento

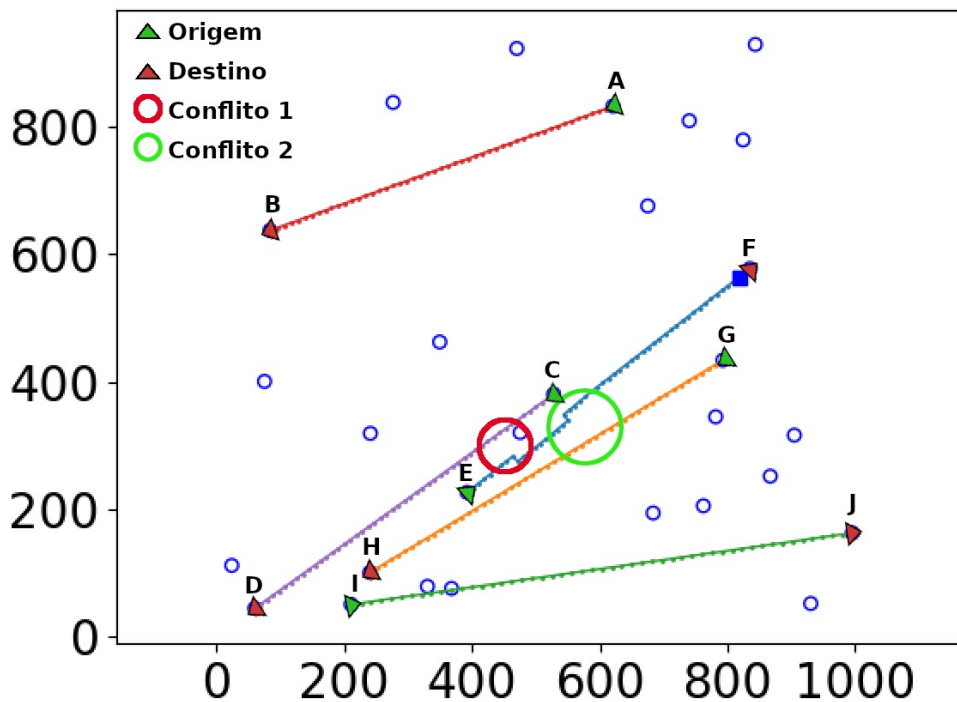


Figura 7.1: Visualização de dois conflitos em um episódio.

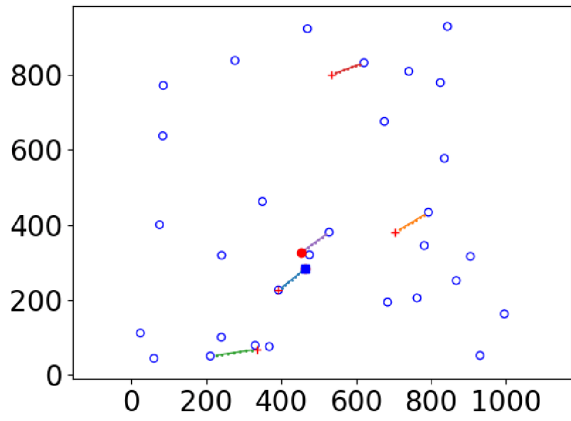
depois que a aeronave azul executou um desvio à direita e logo em seguida fez uma curva à esquerda para retomar a direção do seu destino no ponto F.

De forma semelhante ao conflito 1, a Figura 7.2c mostra o momento em que o conflito 2 entre a aeronave azul e a aeronave laranja é detectado. Por fim a Figura 7.2d retrata o momento logo após a realização do desvio quando a aeronave azul retoma o voo em direção ao ponto F.

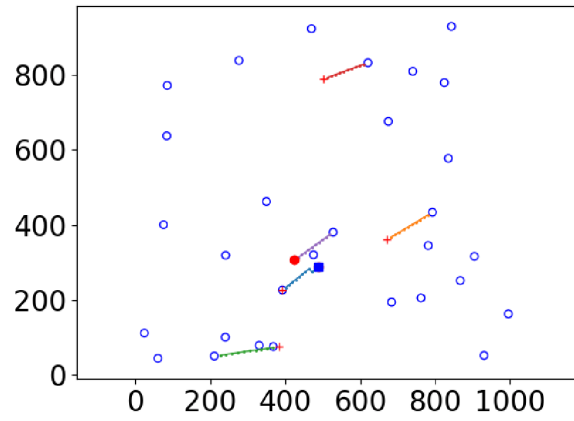
7.3 Teste de Duração do Treinamento

O primeiro teste realizado foi para determinação da duração necessária para que o treinamento atinja seu potencial de resolução de conflitos de acordo com o modelo escolhido. Para isso, o número de episódios aplicado neste teste foi 16.000, pois testes preliminares já indicavam que esse número seria suficiente para visualizar a convergência do modelo. O modelo aplicado neste teste foi o *Double Deep Q learning Network* (DDQN), descrito em 2.4.3. O mesmo teste foi aplicado a outros modelos com resultados semelhantes, por isso os gráficos não serão repetidos neste texto.

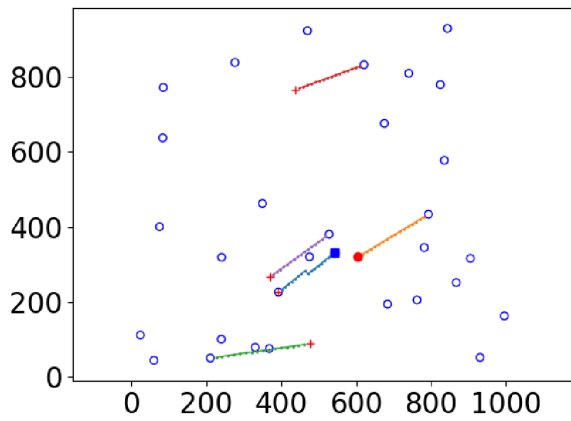
A figura 7.3 mostra uma redução acentuada na quantidade de conflitos nos primeiros mil episódios. Depois a queda continua porém de forma mais suave até por volta do episódio de número 3.000. Percebe-se que após um determinado momento o modelo passa



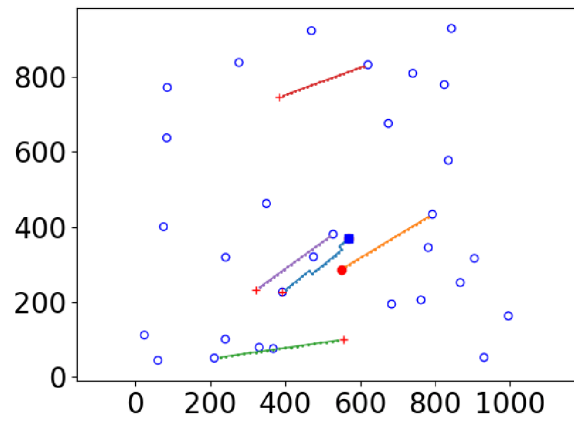
(a) Início do conflito 1



(b) Conflito 1 após o desvio



(c) Início do conflito 2



(d) Conflito 2 após o desvio

Figura 7.2: Detalhes dos conflitos exibidos na Figura 7.1

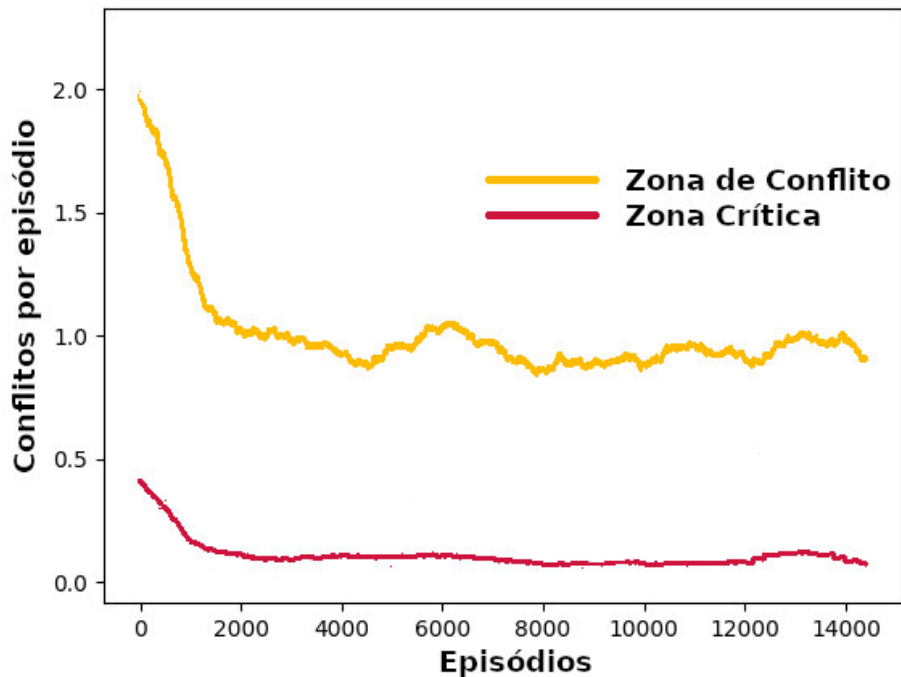


Figura 7.3: Quantidade de episódios até a convergência do modelo.

a apresentar oscilações naturais, por isso os demais testes foram interrompidos após no máximo 10.000 episódios.

7.4 Comparação entre Agentes de DRL

Os testes descritos a seguir foram realizados com o objetivo de comparar o desempenho dos agentes de DRL descritos na seção 2.4 na resolução de conflitos, mais especificamente para comparar a capacidade de evitar o ingresso de intrusos na zona crítica de proteção da aeronave.

Os modelos *Deep Q Learning Network*, *Double Deep Q Learning*, *Dueling Deep Q Learning*, *Proximal Policy Optimization*, *Actor-Critic* e *Asynchronous Advantage Actor-Critic* serão referenciados nesta seção, para simplicidade, apenas como DQN, DDQN, Dueling, PPO, AC e A3C, respectivamente.

Todos os modelos citados acima foram testados em treinamentos com duração de 10.000 episódios. A evolução do número de intrusos que ingressaram na zona crítica de conflito pode ser observada na Figura 7.4. Nota-se que os modelos DQN, DDQN e Dueling tiveram desempenho expressivamente melhor em relação aos demais. O bom resultado

traduz-se numa redução significativa do número de conflitos após 4.000 episódios, quando seria esperado que os modelos atingissem a convergência.

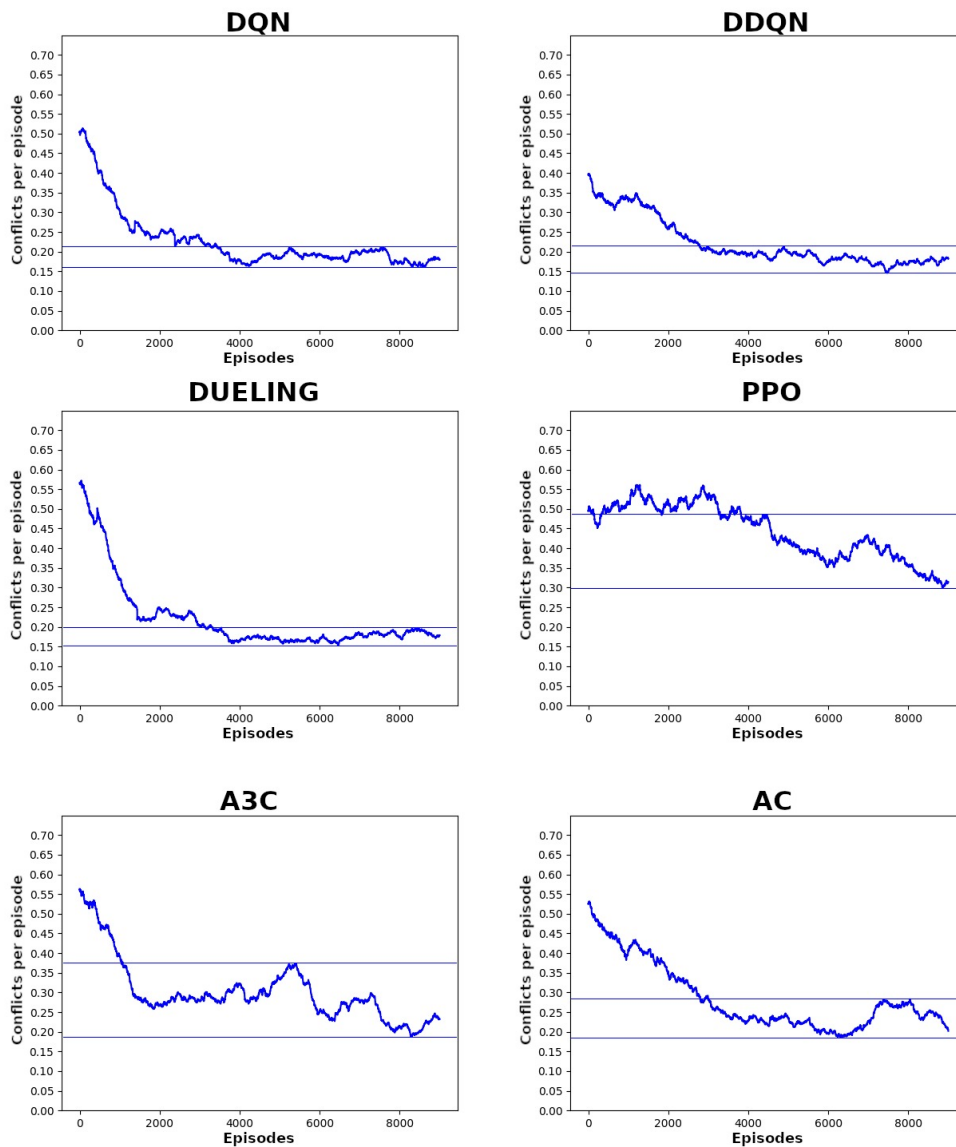


Figura 7.4: Comparação entre os agentes de DRL.

Adicionalmente, uma medida da qualidade desse resultado é a baixa amplitude da oscilação dos valores obtidos após a marca de 4.000 episódios. Essa baixa oscilação pode ser medida com a variável Δ que é o resultado da subtração do valor demarcado pela faixa superior do valor da faixa inferior, mostradas como linhas horizontais na figura 7.4. Um baixo valor Δ indica que o modelo apresenta uma maior consistência entre os episódios, portanto há uma confiabilidade maior no que se pode esperar daquele algoritmo quando

aplicado na resolução de conflitos. Os valores médios após 4.000 episódios e também o valor de Δ para cada modelo podem ser visualizados na tabela 7.2.

Tabela 7.2: Comparação entre Agentes de DRL

Modelo	Média	Δ	% Redução
DQN	0.187	0.054	62,6
DDQN	0.181	0.070	63,8
Dueling	0.176	0.048	64,8
PPO	0.392	0.189	21,6
A3C	0.282	0.188	43,6
AC	0.235	0.098	53,0

Por fim, apresentamos a métrica que representa o percentual de conflitos por episódio. Os gráficos da figura 7.4 mostram no eixo das ordenadas o número de vezes em que o intruso ingressou na zona crítica a cada episódio do treinamento. Como seria esperado, este valor diminui com o passar dos episódios. Tomando 0,5 como valor médio inicial do número de conflitos por episódio, uma característica da configuração do simulador que vale para todos os modelos, podemos calcular o percentual de redução de conflitos que cada um dos modelos é capaz de alcançar. Estes valores são apresentados na última coluna da tabela 7.2.

Como pode ser observado tanto por meio dos gráficos quanto da tabela, o modelo **Dueling** apresentou os melhores resultados em todas as métricas consideradas. Por isso esse foi definido como o modelo de DRL que será utilizado no ISUAM. Além disso, os três modelos baseados em DQN apresentaram melhores números melhores do que os demais modelos. Especialmente quanto à estabilidade, refletida no valor Δ , observa-se que os modelos PPO, A3C e AC apresentaram valores significativamente piores.

7.4.1 Comparação com as Estratégias Alternativas

Foram realizados testes para comparar a evolução do ISUAM e das estratégias alternativas após o treinamento em relação à capacidade de evitar o ingresso de intrusos nas zonas de proteção, ou seja, na zona de conflito e na zona crítica. Nestes gráficos, o ideal é que uma estratégia seja capaz de manter os valores o mais próximo de zero possível.

A figura 7.5 mostra o resultado do teste comparando as estratégias de resolução de conflito alternativas descritas na seção 6.2. Observa-se que a estratégia Free obteve o pior desempenho pois não leva em consideração a existência de intrusos tampouco realiza manobras evasivas. Como foi mencionado, o objetivo da definição dessa estratégia foi estabelecer um limite superior, o qual não deveria ser ultrapassado pelo modelo de ISUAM.

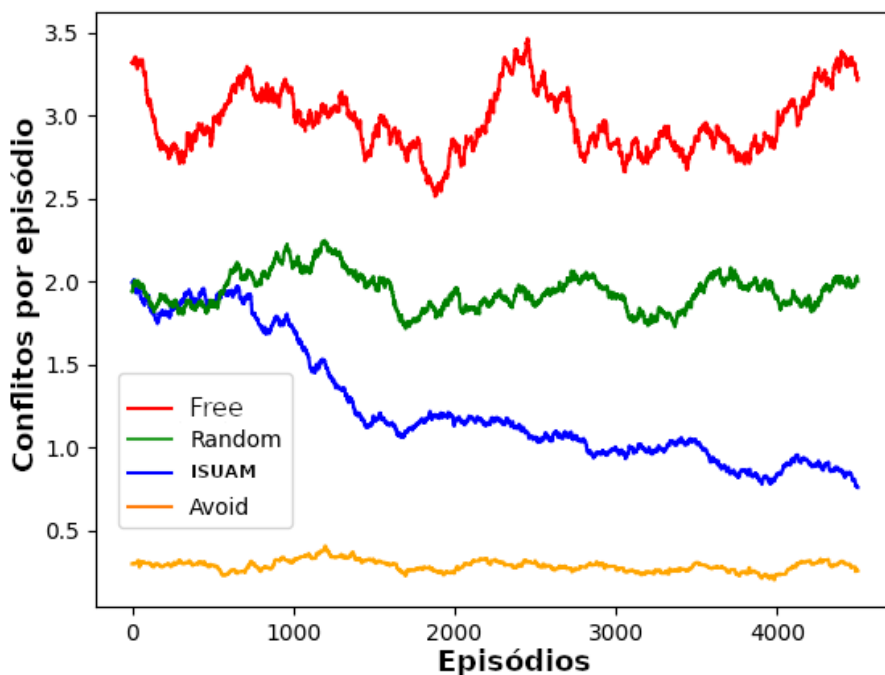


Figura 7.5: Quantidade de intrusos na zona de conflito durante o treinamento.

Por sua vez, a estratégia Avoid obteve o menor valor apresentado no gráfico. De forma análoga à estratégia Free, o objetivo de Avoid é estabelecer um limite inferior, um valor cujo objetivo seria o ideal a ser alcançado pelo ISUAM. Porém veremos em outro teste que os desvios realizados para evitar um conflito por parte desta estratégia fazem com que a aeronave se afaste excessivamente da sua trajetória ideal, prejudicando o desempenho em relação ao tempo de deslocamento e consumo de combustível.

A estratégia Random manteve um valor aproximadamente constante ao redor de 2 conflitos por episódio. Essa estratégia serve como uma linha base a partir da qual o valor do modelo ISUAM iniciará e o seu afastamento da estratégia Random é um indicativo de que o treinamento está evoluindo.

O ISUAM iniciou com um valor ao redor de 2 conflitos por episódio porém este número decaiu com o passar dos episódios, chegando a um valor abaixo de 1 no final do treinamento. Isso indica uma capacidade superior a 50% de evitar o ingresso de intrusos na zona de conflito.

De forma semelhante, a figura 7.6 mostra a evolução das estratégias em relação ao número de intrusos que ingressaram na zona crítica com o passar dos episódios. Os desempenhos das estratégias Avoid, Free e Random são semelhantes à análise feita anteriormente, apresentando apenas valores menores. O ISUAM também iniciou com um valor próximo àquele apresentado pela estratégia Random. Porém a evolução do treinamento fez com que o número caísse a um valor muito próximo daquele apresentado pela

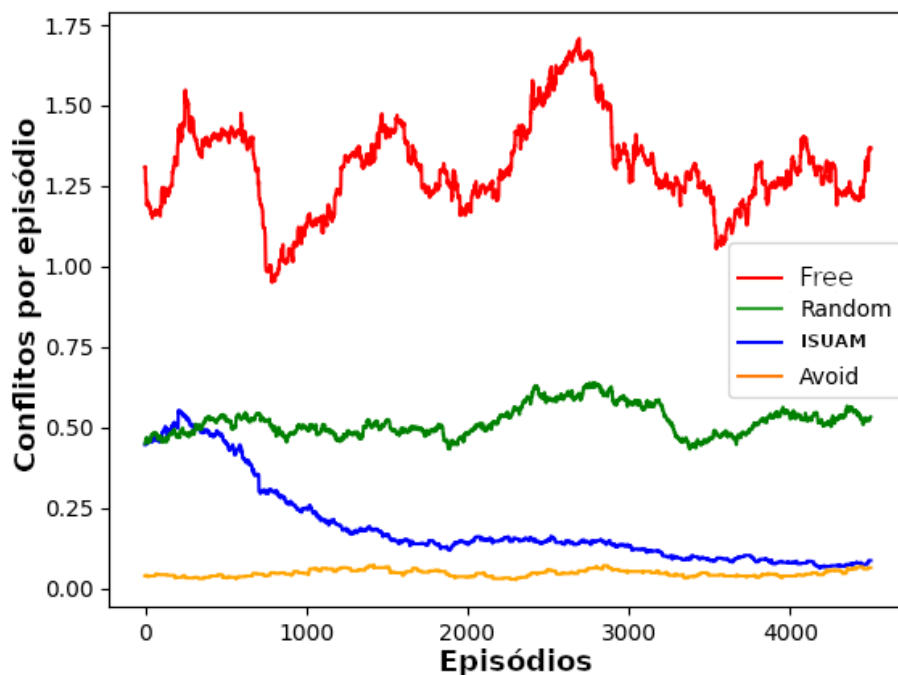


Figura 7.6: Quantidade de intrusos na zona crítica durante o treinamento.

estratégia Avoid, considerado o valor ideal para essa análise.

7.5 Densidade de Tráfego

Os efeitos da densidade de tráfego na capacidade de resolução de conflitos pelo modelo ISUAM também foram investigados. Para essa finalidade, foram utilizados três cenários distintos: um cenário com baixa densidade, composto apenas por três aeronaves em cada episódio; um cenário com densidade média, no qual 7 aeronaves estavam em voo simultâneo; e um cenário com alta densidade, com quatorze voos simultâneos.

O cenário com alta densidade de tráfego apresentou um número maior de conflitos, mesmo após a convergência do modelo. Tal comportamento já era esperado pois a quantidade maior de aeronaves tende a dificultar a resolução de conflitos. Entretanto, no cenário de baixa densidade, estranhamente o número de conflitos não foi menor se comparado ao cenário de média densidade. Por isso o gráfico da simulação de baixa densidade será apresentado com as estratégias alternativas para discussão.

De forma semelhante à discussão ocorrida na seção anterior, no gráfico da figura 7.8 as estratégias Free e Avoid apresentaram, respectivamente, os limites superior e inferior do que poderia ser obtido na simulação em questão. O comportamento inesperado do modelo ISUAM aparece com um aumento inicial no número de conflitos, voltando a diminuir

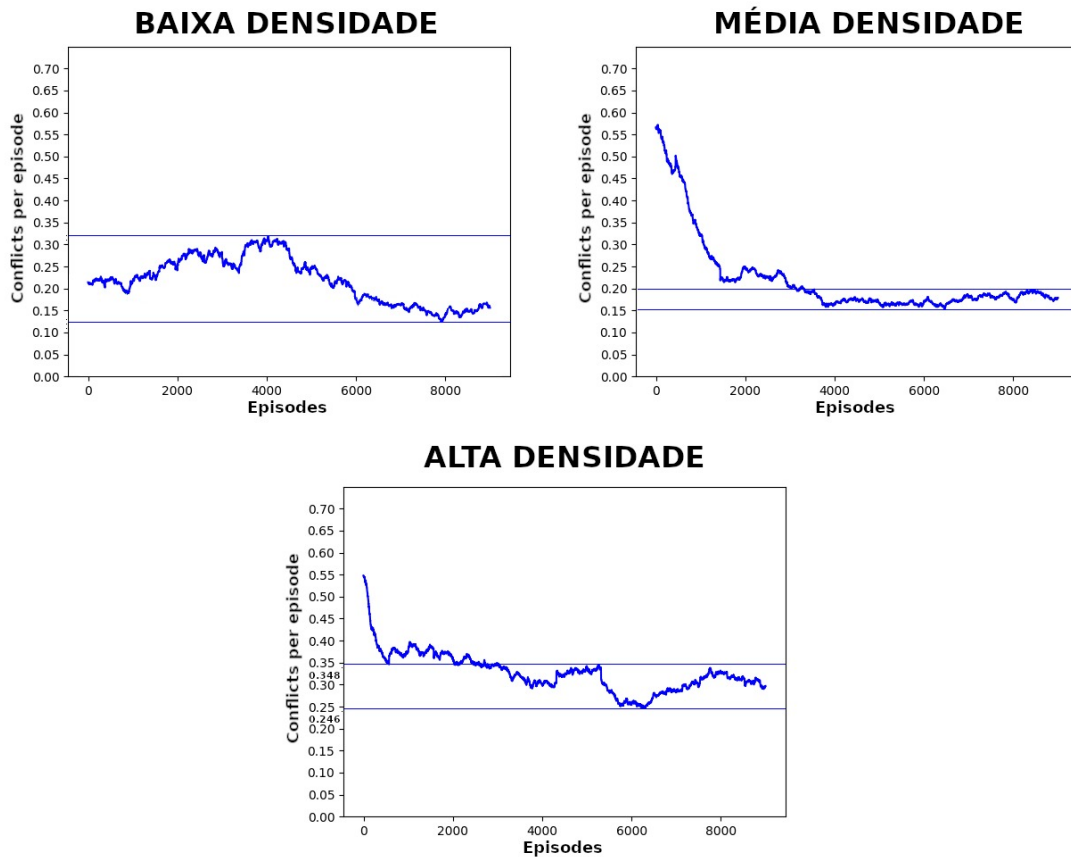


Figura 7.7: Desempenho do modelo ISUAM com variações na densidade de tráfego.

após 4.000 episódios. O valor ultrapassa a linha da estratégia Random e somente passa a ficar menor bem próximo do final da simulação. Outros testes foram realizados com a mesma configuração e apresentaram resultados semelhantes. Portanto, isso indica que tal comportamento é uma característica deste modelo neste cenário. Será necessário um estudo mais aprofundado para buscar compreender a causa deste comportamento.

7.6 Eficiência na Resolução de Conflitos

Uma medida tão importante quanto a capacidade de resolução de conflitos é o comprimento da rota adicional percorrida. Afinal um sistema que evita completamente os desvios mas não permite que a aeronave chegue ao seu destino final não poderá ser utilizado em situações reais. Para avaliar o comprimento adicionado à rota original em virtude dos desvios, foi realizado o teste demonstrado na figura 7.9. Neste gráfico os valores menores indicam um desempenho melhor.

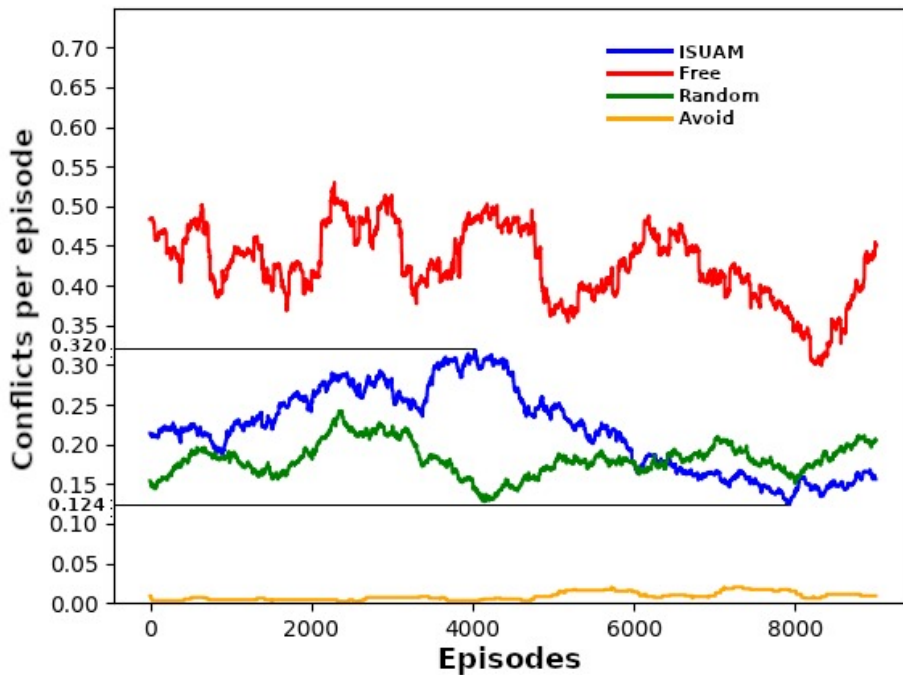


Figura 7.8: Comportamento do modelo ISUAM no cenário de baixa densidade.

Na comparação com as outras estratégias a situação se inverte com relação aos gráficos analisados anteriormente. A estratégia Avoid obteve os maiores valores do gráfico, justamente por realizar um excesso de desvios sem obedecer a qualquer critério. A estratégia Random apresentou um valor com pequena oscilação mas ainda bastante alto se comparado com o ISUAM. Por fim a estratégia Free não é apresentada no gráfico já que seria apenas uma linha horizontal com valor zero, visto que não realiza nenhum desvio.

O modelo ISUAM teve um ligeiro acréscimo após a conclusão do treinamento, considerando que esse momento ocorreu na casa dos 4.000 episódios. Um certo aumento é esperado, afinal a aeronave precisa realizar manobras para poder desviar dos intrusos. Mas o acréscimo não apresenta um valor significativo.

7.7 Discussão

Foram realizados diversos testes tanto para o estabelecimento de parâmetros ideais do simulador quanto para investigar os resultados que os modelos de DRL sob análise são capazes de obter. Conclui-se, a partir dos experimentos, que os modelos sob análise são capazes de chegar à convergência com pouco mais de 4.000 episódios.

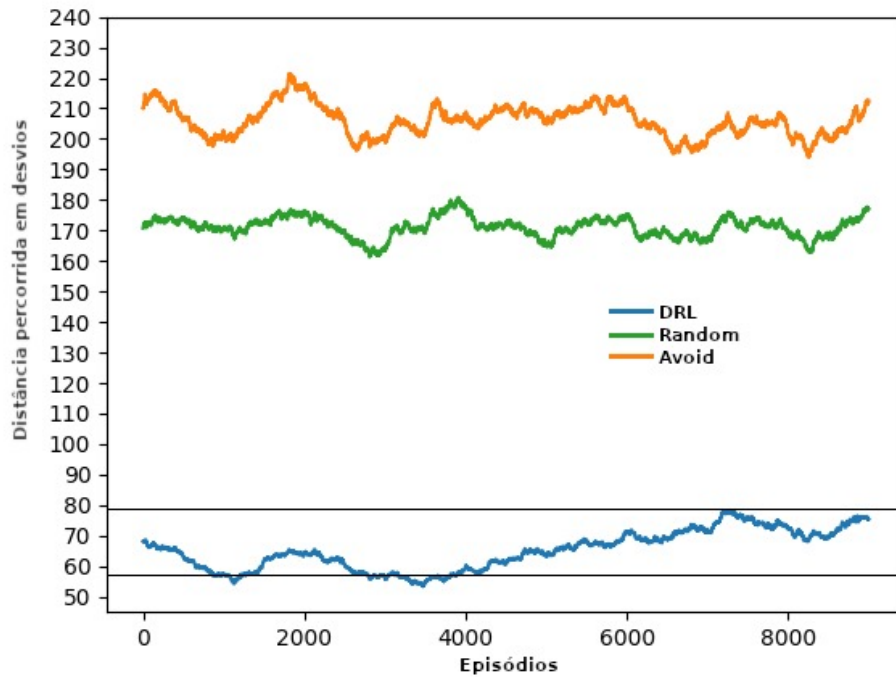


Figura 7.9: Comparação entre as estratégias da distância percorrida no desvio.

O modelo Dueling DQN apresentou melhores resultados entre os modelos analisados, por isso foi adotado como o modelo a ser utilizado no ISUAM. Este modelo obteve o desempenho esperado em relação às estratégias alternativas. Na análise da densidade de tráfego, o desempenho sob alta e média densidade atingiu os valores esperados, porém sob baixa densidade observou-se um comportamento que requer análises mais cuidadosas em trabalhos futuros.

Por fim, a análise da eficiência na resolução dos conflitos, traduzida no caminho adicional percorrido pela aeronave em relação a uma em que não houvesse conflitos mostrou que o modelo ISUAM apresentou ligeiro aumento na distância, porém o valor do acréscimo não foi considerado significativo.

Capítulo 8

Conclusão

A mobilidade aérea urbana apresenta-se como uma alternativa promissora para contribuir na solução dos desafios da locomoção nas grandes cidades. Embora os helicópteros sejam uma opção existente há décadas, a acessibilidade e o alto custo tornam inviável sua utilização por grande parte da população. A chegada das aeronaves com propulsão elétrica pode mudar essa realidade, mas é necessário garantir a segurança do espaço aéreo e desenvolver gerenciadores de trajetórias para garantir a viabilidade do transporte aéreo em larga escala. A UAM não só pode revolucionar o transporte de passageiros como também abrange outras áreas, como logística e emergências médicas. No entanto, é preciso considerar a sustentabilidade e a qualidade de vida das pessoas ao desenvolver essa modalidade de transporte para evitar impactos negativos ao meio ambiente e à saúde pública.

A aprendizagem de reforço profunda é uma técnica de aprendizado de máquina que permite que um agente aprenda a tomar decisões ótimas em um ambiente incerto e complexo. Essa técnica tem sido amplamente utilizada em jogos, robótica e outras aplicações. No problema de gerenciamento de trajetórias, o DRL foi analisado como uma solução para garantir a segurança das operações de aeronaves eVTOL.

A aplicação da aprendizagem de reforço profunda em operações de eVTOL pode ajudar a melhorar a eficiência do gerenciamento de tráfego aéreo. O DRL pode ser usado para treinar os agentes a tomar decisões em tempo real, considerando as condições do ambiente e as interações com outras aeronaves. Esses agentes podem aprender a otimizar rotas e evitar colisões, tornando as operações aéreas mais seguras.

8.1 Contribuições

O presente trabalho visa auxiliar na evolução do campo de pesquisa no qual os métodos de DRL são utilizados para desenvolvimento de sistemas de gerenciamento de trajetórias

de aeronaves eVTOL no UAM.

Entre as contribuições alcançadas, podemos destacar:

- Revisão da literatura para determinar quais trabalhos anteriores já haviam analisado o problema em questão e suas abordagens.
- Modelagem do problema de resolução de conflitos como um problema de aprendizado de reforço no qual o agente de DRL poderá executar ações de modo a obter recompensas quando forem realizados os desvios de forma satisfatórias.
- Implementação de um simulador específico para a tarefa em questão e definição dos melhores parâmetros a serem empregados em cada modelo de DRL.
- Comparação entre diversos modelos de DRL para determinar sua capacidade de evitar conflitos utilizando mudanças de direção.
- Criação de um sistema denominado ISUAM para solução de conflitos sob diversas densidades de tráfego.
- Análise da eficiência das rotas utilizadas para a realização dos desvios.

Assim, foi alcançado o principal objetivo do trabalho com o desenvolvimento de um sistema de gerenciamento de trajetórias de aeronaves eVTOL no ambiente UAM capaz de proporcionar voos livres garantindo a separação entre as aeronaves envolvidas, utilizando ISUAM para oferecer uma navegação segura e eficiente.

8.2 Trabalhos Futuros

Este trabalho investigou apenas a modificação da direção das aeronaves como forma de solucionar conflitos. Outras opções seriam a modificação da altitude de voo e também da velocidade. Trabalhos futuros poderão investigar estas opções de forma isolada ou também a combinação das ações para determinar se essa configuração pode levar a estratégias mais eficientes de solução de conflitos.

A utilização de um simulador customizado permitiu a análise dos resultados apresentados e o preenchimento de algumas das lacunas existentes em relação a outros trabalhos relacionados. Porém a validação desses resultados em um simulador como de código aberto como o BlueSky [72] seria uma contribuição importante para esse campo de estudos.

Como foi visto em 3.4.2, a resolução de conflitos no UAM ocorrerá em várias fases. O sistema proposto neste trabalho visa oferecer uma solução apenas para a fase tática de resolução de conflitos. A integração com sistemas que venham a ser desenvolvidos para a solução de conflitos em outras fases também é um problema que poderá ser estudado no futuro.

Por fim, sugere-se uma análise estatística mais profunda com o objetivo de compreender melhor o resultado obtido na Seção 7.5 em relação ao cenário de baixa densidade.

Referências

- [1] Cohen, Adam P, Susan A Shaheen e Emily M Farrar: *Urban air mobility: History, ecosystem, market potential, and challenges*. IEEE Transactions on Intelligent Transportation Systems, 22(9):6074–6087, 2021. 1, 2
- [2] Gakenheimer, Ralph: *Urban mobility in the developing world*. Transportation Research Part A: Policy and Practice, 33(7-8):671–689, 1999. 1
- [3] Pojani, Dorina e Dominic Stead: *Sustainable urban transport in the developing world: beyond megacities*. Sustainability, 7(6):7784–7805, 2015. 1
- [4] Straubinger, Anna, Raoul Rothfeld, Michael Shamiyeh, Kai Daniel Büchter, Jochen Kaiser e Kay Olaf Plötner: *An overview of current research and developments in urban air mobility—Setting the scene for UAM introduction*. Journal of Air Transport Management, 87:101852, 2020. 2
- [5] Mueller, Eric R, Parmial H Kopardekar e Kenneth H Goodrich: *Enabling airspace integration for high-density on-demand mobility operations*. Em *17th AIAA Aviation Technology, Integration, and Operations Conference*, página 3086, 2017. 2
- [6] Blume, Matti: *Protótipo de eVTOL denominado Volocopter 2X exibido na IAA 2017* In: *WIKIMEDIA COMMONS, a midiateca livre*. Flórida: Wikimedia Foundation, 2014. Disponível sob CC BY-SA 4.0 em: [https://commons.wikimedia.org/wiki/File:Volocopter,IAA_2017,_Frankfurt_\(1Y7A1911\).jpg](https://commons.wikimedia.org/wiki/File:Volocopter,IAA_2017,_Frankfurt_(1Y7A1911).jpg). 3
- [7] Edwards, Thomas e George Price: *eVTOL passenger acceptance*. Relatório Técnico, 2020. 3
- [8] Thippavong, David P, Rafael Apaza, Bryan Barmore, Vernol Battiste, Barbara Burian, Quang Dao, Michael Feary, Susie Go, Kenneth H Goodrich, Jeffrey Homola e others: *Urban air mobility airspace integration concepts and considerations*. Em *2018 Aviation Technology, Integration, and Operations Conference*, página 3676, 2018. 4
- [9] Bauranov, Aleksandar e Jasenka Rakas: *Designing airspace for urban air mobility: A review of concepts and approaches*. Progress in Aerospace Sciences, 125:100726, 2021. 5
- [10] Schultz, Robert, Donald Shaner, Yiyuan Zhao, Robert Schultz, Donald Shaner e Yiyuan Zhao: *Free flight concept*. Em *Guidance, Navigation, and Control Conference*, página 3677, 1997. 5, 38

- [11] Yang, Xuxi e Peng Wei: *Scalable multi-agent computational guidance with separation assurance for autonomous urban air mobility*. *Journal of Guidance, Control, and Dynamics*, 43(8):1473–1486, 2020. 5, 9, 10
- [12] François-Lavet, Vincent, Peter Henderson, Riashat Islam, Marc G Bellemare, Joelle Pineau e others: *An introduction to deep reinforcement learning*. *Foundations and Trends® in Machine Learning*, 11(3-4):219–354, 2018. 5, 13
- [13] Justesen, Niels, Ruben Rodriguez Torrado, Philip Bontrager, Ahmed Khalifa, Julian Togelius e Sebastian Risi: *Illuminating generalization in deep reinforcement learning through procedural level generation*. *arXiv preprint arXiv:1806.10729*, 2018. 5
- [14] Feinberg, Eugene A e Adam Shwartz: *Handbook of Markov decision processes: methods and applications*, volume 40. Springer Science & Business Media, 2012. 9
- [15] Haykin, Simon: *Neural networks and learning machines, 3/E*. Pearson Education India, 2009. 11
- [16] Abraham, Ajith: *Artificial neural networks*. *Handbook of measuring system design*, 2005. 11
- [17] Mayranna: *Perceptron*. Flórida: *Wikimedia Foundation, 2014. Disponível sob CC BY-SA 3.0 em: https://commons.wikimedia.org/wiki/File:Perceptron_moj.png*. 11
- [18] Glosser.ca: *Rede neural artificial*. Flórida: *Wikimedia Foundation, 2014. Disponível sob CC BY-SA 3.0 em: https://commons.wikimedia.org/wiki/File:Colored_neural_network.svg*. 11
- [19] Krogh, Anders: *What are artificial neural networks?* *Nature biotechnology*, 26(2):195–197, 2008. 12
- [20] Mitchell, Tom M e Tom M Mitchell: *Machine learning*, volume 1. McGraw-hill New York, 1997. 12, 13, 21
- [21] O’Shea, Keiron e Ryan Nash: *An introduction to convolutional neural networks*. *arXiv preprint arXiv:1511.08458*, 2015. 14, 15
- [22] Cohen, Gregory, Saeed Afshar, Jonathan Tapson e Andre Van Schaik: *EMNIST: Extending MNIST to handwritten letters*. Em *2017 international joint conference on neural networks (IJCNN)*, páginas 2921–2926, 2017. 14
- [23] Aphex34: *Rede neural convolucional totalmente conectada*. Flórida: *Wikimedia Foundation, 2015. Disponível sob CC BY-SA 4.0 em: https://commons.wikimedia.org/wiki/File:Typical_cnn.png*. 15
- [24] Mehryar Mohri, AfshinRostamizadeh e A Talwalka: *Foundations of machine learning*, 2012. 15, 16

- [25] Pamplona, Daniel Alberto, Li Weigang, Alexandre Gomes de Barros, Elcio Hideiti Shiguemori e Claudio Jorge Pinto Alves: *Supervised neural network with multilevel input layers for predicting of air traffic delays*. Em *2018 International Joint Conference on Neural Networks (IJCNN)*, páginas 1–6, 2018. 16
- [26] Haldorai, A, A Ramu e M Suriya: *Organization internet of things (IoTs): supervised, unsupervised, and reinforcement learning*. Em *Business Intelligence for Enterprise Internet of Things*, páginas 27–53. Springer, 2020. 16
- [27] Li, Yuxi: *Deep reinforcement learning: An overview*. arXiv preprint arXiv:1701.07274, 2017. 17
- [28] Jiang, Tammy, Jaimie L Gradus e Anthony J Rosellini: *Supervised machine learning: a brief primer*. *Behavior Therapy*, 51(5):675–687, 2020. 17
- [29] Conde Rocha Murca, Mayara, Richard DeLaura, R John Hansman, Richard Jordan, Tom Reynolds e Hamsa Balakrishnan: *Trajectory clustering and classification for characterization of air traffic flows*. Em *16th AIAA Aviation Technology, Integration, and Operations Conference*, página 3760, 2016. 18
- [30] Sutton, Richard S e Andrew G Barto: *Reinforcement learning: An introduction*. MIT press, 2018. 18, 19
- [31] EBatlleP: *Cliclo do aprendizado de reforço In: WIKIMEDIA COMMONS, a midiateca livre. Flórida: Wikimedia Foundation, 2014. Disponível sob CC BY-SA 4.0 em: https://commons.wikimedia.org/wiki/File:Markov_diagram_v2.svg*. 19
- [32] Dewey, Daniel: *Reinforcement learning and the reward engineering principle*. Em *2014 AAAI Spring Symposium Series*, 2014. 20
- [33] Jin, Chi, Akshay Krishnamurthy, Max Simchowitz e Tiancheng Yu: *Reward-free exploration for reinforcement learning*. Em *International Conference on Machine Learning*, páginas 4870–4879, 2020. 20
- [34] Crespo, Antonio Marcio Ferreira, Li Weigang e Alexandre Gomes de Barros: *Reinforcement learning agents to tactical air traffic flow management*. *International Journal of Aviation Management*, 1(3):145–161, 2012. 20, 21
- [35] Arulkumaran, Kai, Marc Peter Deisenroth, Miles Brundage e Anil Anthony Bharath: *Deep reinforcement learning: A brief survey*. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017. 21
- [36] Mousavi, Seyed Sajad, Michael Schukat e Enda Howley: *Deep reinforcement learning: an overview*. Em *Proceedings of SAI Intelligent Systems Conference*, páginas 426–440, 2016. 21, 23
- [37] Jang, Beakcheol, Myeonghwi Kim, Gaspard Harerimana e Jong Wook Kim: *Q-learning algorithms: A comprehensive classification and applications*. *IEEE access*, 7:133653–133667, 2019. 22, 24

- [38] Mnih, Volodymyr, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski e others: *Human-level control through deep reinforcement learning*. *nature*, 518(7540):529–533, 2015. 22
- [39] Van Hasselt, Hado, Arthur Guez e David Silver: *Deep Reinforcement Learning with Double Q-Learning*. Em *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016. 24
- [40] Sewak, Mohit: *Deep Q-Network (DQN), Double DQN, and Dueling DQN*. Em *Deep Reinforcement Learning*, páginas 95–108. Springer, 2019. 25
- [41] Wang, Ziyu, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot e Nando Freitas: *Dueling network architectures for deep reinforcement learning*. Em *International conference on machine learning*, páginas 1995–2003, 2016. 25
- [42] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford e Oleg Klimov: *Proximal policy optimization algorithms*. arXiv preprint arXiv:1707.06347, 2017. 26
- [43] Kiran, B Ravi, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A Al Sallab, Senthil Yogamani e Patrick Pérez: *Deep reinforcement learning for autonomous driving: A survey*. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, 2021. 26
- [44] Grondman, Ivo, Lucian Busoniu, Gabriel A D Lopes e Robert Babuska: *A survey of actor-critic reinforcement learning: Standard and natural policy gradients*. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):1291–1307, 2012. 26, 27
- [45] Mnih, Volodymyr, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver e Koray Kavukcuoglu: *Asynchronous methods for deep reinforcement learning*. Em *International conference on machine learning*, páginas 1928–1937, 2016. 27
- [46] Babaeizadeh, Mohammad, Iuri Frosio, Stephen Tyree, Jason Clemons e Jan Kautz: *Reinforcement learning through asynchronous advantage actor-critic on a gpu*. arXiv preprint arXiv:1611.06256, 2016. 28
- [47] Abeyratne, Ruwantissa e Ruwantissa Abeyratne: *The Nature and Role of ICAO. Regulation of Air Transport: The Slumbering Sentinels*, páginas 1–25, 2014. 29
- [48] Mackenzie, David: *ICAO: A history of the international civil aviation organization*. University of Toronto Press, 2010. 30
- [49] ICAO: *Annex 2 - Rules of the Air*. Relatório Técnico, Montreal, 2005. 30
- [50] ICAO: *Air Traffic Services: Annex 11 to the Convention on International Civil Aviation*. International Civil Aviation Organ, 2001. 32
- [51] DECEA: *ICA 100-37 Serviços de Tráfego Aéreo*, 2020. 32, 36

- [52] Jardin, Matt: *Grid-based strategic air traffic conflict detection*. Em *AIAA Guidance, Navigation, and Control Conference and Exhibit*, página 5826, 2005. 36
- [53] Thomas, Lisa C e Christopher D Wickens: *Display dimensionality and conflict geometry effects on maneuver preferences for resolving in-flight conflicts*. *Human factors*, 50(4):576–588, 2008. 37
- [54] Ribeiro, Marta, Joost Ellerbroek e Jacco Hoekstra: *Review of conflict resolution methods for manned and unmanned aviation*. *Aerospace*, 7(6):79, 2020. 38, 43, 45, 46
- [55] Goodrich, Kenneth H e Colin R Theodore: *Description of the NASA urban air mobility maturity level (UML) scale*. Em *AIAA Scitech 2021 Forum*, página 1627, 2021. 38
- [56] Weigang, Li e Cristiano P Garcia: *Air Traffic Flow and Capacity Management*. 2021. 43
- [57] Ball, Michael O, Chien Yu Chen, Robert Hoffman e Thomas Vossen: *Collaborative decision making in air traffic management: Current and future research directions*. Springer, 2001. 43
- [58] Manfredi, Guido e Yannick Jestin: *An introduction to ACAS Xu and the challenges ahead*. Em *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*, páginas 1–9, 2016. 44
- [59] Xue, Min: *Urban Air Mobility Conflict Resolution: Centralized or Decentralized?* Em *AIAA Aviation 2020 Forum*, página 3192, 2020. 46
- [60] Hoekstra, Jacco M, Ronald N H W van Gent e Rob C J Ruijgrok: *Designing for safety: the ‘free flight’ air traffic management concept*. *Reliability Engineering & System Safety*, 75(2):215–232, 2002. 46, 47, 54
- [61] Balasooriyan, Suthes: *Multi-aircraft conflict resolution using velocity obstacles*, outubro 2017. 47, 48, 54
- [62] Paielli, Russell A: *Modeling maneuver dynamics in air traffic conflict resolution*. *Journal of guidance, control, and dynamics*, 26(3):407–415, 2003. 47, 48, 54
- [63] Van Dam, Stijn, Max Mulder e René Paassen: *The use of intent information in an airborne self-separation assistance display design*. Em *AIAA Guidance, Navigation, and Control Conference*, página 5745, 2009. 48, 49, 54
- [64] Velasco, Gustavo Adrián Mercado, Clark Borst, Joost Ellerbroek, M M Van Paassen e Max Mulder: *The use of intent information in conflict detection and resolution models based on dynamic velocity obstacles*. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):2297–2302, 2015. 49, 50, 54
- [65] Ortner, Philipp, Raphael Steinhöfler, Erich Leitgeb e Holger Flühr: *Augmented Air Traffic Control System—Artificial Intelligence as Digital Assistance System to Predict Air Traffic Conflicts*. *AI*, 3(3):623–644, 2022. 50

- [66] Xie, Hua, Minghua Zhang, Jiaming Ge, Xinfang Dong e Haiyan Chen: *Learning air traffic as images: a deep convolutional neural network for airspace operation complexity evaluation*. Complexity, 2021, 2021. 50
- [67] Naessens, Herbert, Thomas Philip, Marcin Piatek, Kristof Schippers e Robert Parys: *Predicting flight routes with a Deep Neural Network in the operational Air Traffic Flow and Capacity Management system*. EUROCONTROL Maastricht Upper Area Control Centre, Maastricht Airport, The Netherlands, Tech. Rep, 2017. 50
- [68] Abadi, Martin, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard e others: *TensorFlow: a system for Large-Scale machine learning*. Em *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, páginas 265–283, 2016. 50, 63
- [69] Brittain, Marc e Peng Wei: *Autonomous aircraft sequencing and separation with hierarchical deep reinforcement learning*. Em *Proceedings of the International Conference for Research in Air Transportation*, 2018. 50, 51, 54, 55
- [70] Brittain, Marc e Peng Wei: *Autonomous Air Traffic Controller: A Deep Multi-Agent Reinforcement Learning Approach*. arXiv preprint arXiv:1905.01303, 2019. 51, 52, 54, 55
- [71] Brittain, Marc William: *Learning-based decision making for safe and scalable autonomous separation assurance*. Tese de Doutorado, Iowa State University, 2021. 51, 54, 55
- [72] Hoekstra, Jacco M e Joost Ellerbroek: *Bluesky ATC simulator project: an open data and open source approach*. Em *Proceedings of the 7th international conference on research in air transportation*, volume 131, página 132, 2016. 51, 54, 63, 81
- [73] Pham, Duc Thinh, Ngoc Phu Tran, Sameer Alam, Vu Duong e Daniel Delahaye: *A Machine Learning Approach for Conflict Resolution in Dense Traffic Scenarios with Uncertainties*. ATM 2019, 13th USA/Europe Air Traffic Management Research and Development Seminar, 2019. 52, 54, 55
- [74] Li, Sheng, Maxim Egorov e Mykel Kochenderfer: *Optimizing collision avoidance in dense airspace using deep reinforcement learning*. arXiv preprint arXiv:1912.10146, 2019. 53, 54, 55
- [75] Ribeiro, Marta, Joost Ellerbroek e Jacco Hoekstra: *Improving Algorithm Conflict Resolution Manoeuvres with Reinforcement Learning*. Aerospace, 9(12):847, 2022. 53, 54, 55
- [76] Nunes, Tiago Miguel Monteiro, Clark Borst, Erik Jan van Kampen, Brian Hilburn e Carl Westin: *Human-interpretable input for Machine Learning in Tactical Air Traffic Control*. 2021. 54
- [77] Cruciol, Leonardo L B V, Antonio C de Arruda Jr, Li Weigang, Leihong Li e Antonio M F Crespo: *Reward functions for learning to control in air traffic flow management*. Transportation Research Part C: Emerging Technologies, 35:141–155, 2013. 58

[78] Yann Berthelot: *AI learns to fly*, 2020. http://bit.ly/Berthelot_AI. 63

[79] Garcia, Cristiano P: *DRL_UAM*, 2023. https://github.com/cpg100/DRL_UAM. 63

Apêndice A

Produção Científica

4D Trajectory Conflict Detection and Resolution Using Decision Tree Pruning Method

Qualis Periódico B2.

Monteiro, L. B., Ribeiro, V. F., Garcia, C. P., Rocha Filho, G. P., Weigang, L.

The aviation community develops Trajectory Based Operations (TBO) as an advancement in Air Traffic Management (ATM). There is still the need for an efficient scheme to present the trajectories, manage their associated data, and further detect and resolve the conflicts (CDR) that should eventually occur. In this research, we develop a CDR framework for managing predicted 4-Dimensional Trajectory (4DT). Using Not Only SQL (NoSQL) database (Cassandra and MongoDB), the 4D trajectories of related routes are presented, and the possible conflicts are detected using the strategy of Computing in NoSQL Database. Compared with other conflict detection algorithms, usually by the pairwise method with $O(n^2)$ at least, the proposed Decision Tree Pruning Method (DTPM) effectively treats massive data sets. The 4DT data are collected by Trajectory Predictor (TP) concerning 58% of the whole Brazilian air traffic. The comparison results between Cassandra and MongoDB from the case studies show the effectiveness of the proposed methods for conflict detection. In addition, we prove that the conflict resolution approach is viable for application in real scenarios, finding near-optimal solutions for the conflicts identified by the framework. Finally, we also demonstrated the development of sustainable artificial intelligence in intelligent air transportation to improve safety in air traffic management.

IEEE Latin America Transactions, 21(2), 277–287.

Air Traffic Flow and Capacity Management

Li Weigang, Cristiano P Garcia

Air traffic has been increasing in the past years, and the predictions indicate a more significant growth in the future. On the other side, the infrastructure required to provide a safe, orderly, and expeditious flow of traffic does not grow as fast because it is costly. The high price tag encourages the authorities to make sure the existing infrastructure is optimized as much as possible before making any investment to increase capacity. Air Traffic Flow and Capacity Management (ATFCM) is part of Air Traffic Management (ATM) that seeks to balance the increasing demand with a limited capacity. Demand and capacity are on opposite ends of the scale, but proper management of both is a necessary step to provide a quality service for all airspace users.

International Encyclopedia of Transportation. vol. 4, pp. 380-389

Metodologias para Resolução de Conflitos entre Aeronaves Baseadas em Game Tree

Lucas B. Monteiro, Cristiano P. Garcia, Geraldo P. Rocha Filho, Li Weigang

Nos últimos anos, houve um progresso significativo no desenvolvimento de modelos relacionados ao gerenciamento de tráfego aéreo (ATM, do inglês Air Traffic Management). Considerando vários tipos de eventos envolvidos em um voo (precisão dos sensores, incertezas climáticas e meteorológicas, mudanças nas trajetórias das aeronaves, etc.), muitos esforços ainda precisam ser empregados no desenvolvimento de métodos que possam ser aplicados ao ambiente real com máxima precisão. Todos esses métodos precisam lidar com uma grande quantidade de dados. Nesse contexto, alguns pesquisadores têm buscado diferentes metodologias de inteligência artificial para melhorar a tomada de decisão no transporte aéreo. Esta pesquisa tem como objetivo propor dois métodos diferentes de resolução de conflitos baseados em game tree: Busca em Profundidade (DFS, do inglês Depth-First Search) e Alfa-Beta. Serão apresentados os conceitos básicos desses métodos, bem como a modelagem necessária para sua aplicação. Por fim, os resultados das simulações envolvendo diferentes combinações de aeronaves serão analisados, como forma de indicar a possibilidade de aplicação de cada método estudado.

Sitraer 2022.