



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

APRENDIZAGEM POR REFORÇO APLICADA AO GERENCIAMENTO TÁTICO DO FLUXO DE TRÁFEGO AÉREO

Antonio Márcio Ferreira Crespo

Dissertação apresentada como requisito parcial
para conclusão do Mestrado em Informática

Orientador
Prof. Dr. Li Weigang

Brasília, 26 de outubro de 2010

Universidade de Brasília – UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Mestrado em Informática

Coordenador: Prof. Dr. Maurício Ayala-Rincon

Banca examinadora composta por:

Prof. Dr. Li Weigang (Orientador) – CIC/UnB
Prof. Dr. Antonio Pedro Timoszczuk – Atech
Prof.^a Dr.^a Alba Cristina Magalhães A. de Melo – CIC/UnB

CIP – Catalogação Internacional na Publicação

Crespo, Antonio Márcio Ferreira.

APRENDIZAGEM POR REFORÇO APLICADA AO GERENCIAMENTO TÁTICO DO FLUXO DE TRÁFEGO AÉREO / Antonio Márcio Ferreira Crespo. Brasília : UnB, 2010.
119 p. : il. ; 29,5 cm.

Tese (Mestre) – Universidade de Brasília, Brasília, 2010.

1. Aprendizagem por Reforço,
2. Q-learning,
3. Apoio à Decisão,
4. Gerenciamento de Fluxo de Tráfego Aéreo (ATFM)

CDU 004

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro – Asa Norte
CEP 70910-900
Brasília – DF – Brasil

Dedicatória

Este trabalho é dedicado àqueles que sempre foram os maiores incentivadores da minha carreira. Às pessoas que me ensinaram a vencer obstáculos, a nunca desistir...a sempre seguir em frente. Àqueles que viveram para a família, estruturando-a, responsáveis pelo meu sucesso e pelo sucesso dos meus irmãos. A vocês, meus Pais, eu dedico mais esta vitória, que como todas as outras, nós conquistamos juntos. Fiquem com Deus...nós os amamos demais!

Agradecimentos

Agradeço ao meu orientador e grande amigo, *Prof. Dr. Li Weigang*, pela presença constante, pela confiança demonstrada ao longo desses anos de trabalho conjunto e principalmente pela paciência e apoio nos momentos difíceis...

Agradeço aos meus Comandantes e Chefes no CINDACTA I, pelo apoio irrestrito, pela confiança depositada e pelo constante incentivo na busca desse grande objetivo. *Maj Brig Ar Rodrigues Filho, Brig Ar Aquino, Brig Ar Gonçalves, Cel Av Rivera, Cel Av Raulino e Cel Av Belchior*, muito obrigado por tudo...

Por intermédio do *Ten Cel Av Vital* e do *Ten Cel Av Bertolino*, agradeço aos companheiros do CGNA, pelo incondicional apoio, pelas valiosas informações disponibilizadas ao longo de todo o projeto, bem como pela forma acolhedora como sempre fui recebido nesse Centro de excelência...

Agradeço à *Profa. Alba* e aos demais Professores do PPGI, pelos valiosos ensinamentos compartilhados, bem como pelo constante apoio e compreensão a mim dispensados...

Agradeço aos colegas do TRANSLAB, companheiros nos desafios e apreensões que permearam nosso projeto. *Cícero, Júnior, Alessandro, Victor, Leonardo e Bueno*, obrigado pela convivência amistosa e profícua...

Agradeço aos colegas do CIC pelo constante apoio, e em especial à *Rosa* e sua equipe, o meu muito obrigado...

Agradeço à Atech, na pessoa do *Prof. Dr. Antonio Pedro Timoszczuk*, pelo inestimável apoio prestado ao Translab.

Agradeço aos meus queridos filhos, *Thiago e Bruno*, pela compreensão, pois sei que nesses últimos dois anos não brincamos muito; prometo que vou tentar atender os pedidos de vocês e trabalhar um pouco menos daqui para frente!

Agradeço a minha fiel escudeira, companheira de todas as horas e amada esposa, *Débora*, por estar sempre ao meu lado, apoiando, incentivando, mesmo naqueles momentos em que eu bebia algoritmos, respirava modelos e sonhava com a dissertação. Obrigado, te amo!

Finalmente agradeço a *Deus*, por ter me dado forças para continuar mesmo quando a caminhada parecia impossível. Não foram poucos os momentos difíceis, mas em todos eles eu senti a tua presença ao meu lado.

Resumo

O Gerenciamento do Tráfego Aéreo (ATM) se define como um domínio amplo constituído por atividades complexas. No Brasil, o Centro de Gerenciamento da Navegação Aérea (CGNA) é o responsável pelo desempenho do ATFM, cabendo-lhe monitorar e gerenciar 46 setores de controle de tráfego aéreo, 39 áreas de aproximação (TMA), cada qual com pelo menos um aeroporto, além de aproximadamente 5.000 movimentos aéreos diários. Neste contexto emerge a problemática objeto do presente trabalho, qual seja, a inexistência de um sistema de apoio à decisão que auxilie os gerentes de fluxo de tráfego a lidar com o conjunto de variáveis, inerentes ao cenário ATFM, e os auxilie a tomar decisões (adotar medidas de controle de fluxo) na eventualidade da ocorrência de desbalanceamento na relação capacidade X demanda. Este trabalho, portanto, apresenta um protótipo de módulo de apoio à decisão baseado em aprendizagem por reforço que tem como saída sugestões de medidas de controle de fluxo de tráfego geradas a partir de cenários aéreos com pontos de saturação. O módulo desenvolvido foi concebido como um agente inteligente, baseado em aprendizagem, que utiliza o algoritmo Q-learning. O desempenho do protótipo é avaliado por intermédio de um estudo de caso modelado a partir das demandas reais de tráfego verificadas em datas determinadas. O estudo de caso permitiu estabelecer uma análise comparativa entre o desempenho do módulo dadas duas condições. Na primeira condição, o agente atuou a partir de um conjunto de ações que não incorporava a experiência do agente humano (gerente de fluxo de tráfego). Num segundo momento, o agente utilizou um conjunto de ações que incorporava medidas de controle de tráfego adotadas pelos gerentes de fluxo de tráfego em cenários reais. Para a composição de tal conjunto de ações, foram pesquisados todos os registros de ações tomadas entre janeiro de 2008 e março de 2010. O critério básico de avaliação foi a convergência do Algoritmo Q-learning. Os resultados obtidos no estudo de caso são promissores, na medida em que o comportamento dos agentes concebidos indicam que o Algoritmo Q-learning convergiu de forma satisfatória, decorrendo que os protótipos geraram ações que contribuíram de forma efetiva para a redução das saturações nos cenários aéreos testados. Desta feita, com o desenvolvimento dos referidos protótipos, foi possível demonstrar a aplicabilidade de formalismos da Inteligência Artificial na busca de soluções nacionais, para a Problemática ATFM, inteiramente concebidas segundo as características e necessidades do país.

Palavras-chave: Aprendizagem por Reforço, Q-learning, Apoio à Decisão, Gerenciamento de Fluxo de Tráfego Aéreo (ATFM)

Abstract

Air Traffic Management (ATM) is defined as a broad area consisting of complex Activities. ATFM occurs in three different contexts: strategic, pre-tactical and tactical. The tactical ATFM is an activity that aims to maintain an optimum balance between capacity (airspace and airports) and demand (flight intentions), considering to adopt all possible measures in a period of up to two hours prior to the scheduled takeoff time. The Centre for Management of Air Navigation (CGNA) is responsible for the performance of ATFM, and with the daily work of monitoring and managing Brazilian air space, which consists of 46 sectors of air traffic control, 39 areas of approximation (TMA), each with at least one airport, and approximately 5,000 aircraft movements daily. In this context emerges the problematic object of the present work, namely the lack of a decision support system to help the traffic flow managers to deal with all the variables inherent in the scenario ATFM, and help them make decisions (adopt measures to control flow) in the event of imbalance in the ratio capacity X demand. This dissertation therefore presents a prototype of the decision support system based on Artificial Intelligence method, especially, reinforcement learning that has as output measures (suggestions to air traffic flow management) generated from scenarios with saturation points. The developed module was designed as an intelligent agent using the Q-learning algorithm. The performance of the prototype is evaluated through a case study modeled from the real demands of traffic on certain dates. The case study allowed to establish a comparative analysis between the module performance given two conditions. In the first condition, the agent acted from a set of actions produced without the participation of the human agent (manager of traffic flow). Secondly, the agent used a set of actions that incorporated measures adopted by the managers of traffic flow in real scenarios. For the composition of such a set of actions, all records of actions taken between January 2008 and March 2010 were considered. The basic criterion of evaluation was the convergence of Q-learning algorithm. The results showed, as expected, that complex congestion (involving multiple saturation points and of spatially dispersed flow control measures) are handled more efficiently by the computational agent (module). This research demonstrated the feasibility of using reinforcement learning to develop a decision support system applied to tactical ATFM considering Brazilian national scene.

Keywords: Reinforcement Learning, Q-learning, Decision Making Support, Air Traffic Flow Management (ATFM)

Sumário

1	INTRODUÇÃO	11
1.1	Motivação	11
1.2	Objetivo	13
1.3	Metodologia	14
1.3.1	Metodologia de Pesquisa e Implementação	15
1.4	Organização do Trabalho	17
2	GERENCIAMENTO DO TRÁFEGO AÉREO	19
2.1	Conceitos Fundamentais	19
2.1.1	Planejamento Estratégico	20
2.1.2	Planejamento Pré-tático	20
2.1.3	Operações Táticas	21
2.1.4	Capacidade ATC	21
2.1.5	Complexidade do Espaço Aéreo (Complexidade ATC ou Densidade Dinâmica)	22
2.2	O SISCEAB	23
2.2.1	Estrutura do Espaço Aéreo	24
2.3	ATFM no Brasil	25
2.4	Problemática ATFM nas Operações Táticas	28
3	TÉCNICAS COMPUTACIONAIS APLICADAS AO ATFM	33
3.1	Abordagem com Emprego de Programação Linear	35
3.2	Abordagem Empregando Programação Dinâmica e Teoria de Grafos	36
3.3	Abordagem Empregando Programação Inteira e Teoria de Grafos	37
3.4	Abordagem Empregando Técnicas Multi-agente	39
3.5	Abordagem Empregando Aprendizagem por Reforço	40
3.6	Estado da Arte	44
4	FUNDAMENTAÇÃO TEÓRICA	47
4.1	Processo Decisório	47
4.1.1	Modelo de Processamento de Informação para Tomada de Decisão	48

4.1.2	Qualificação da Decisão	48
4.1.3	Diagnóstico e Detecção da Situação	49
4.1.4	Escolha da Ação	50
4.2	Teoria da Decisão	51
4.2.1	Teoria da Utilidade e Tomada de Decisão Simples	51
4.2.2	Tomada de Decisão Complexa (Decisões Seqüenciais)	54
4.3	Processo Decisório de Markov	56
4.4	Aprendizagem por Reforço	57
4.4.1	Algoritmo Q-learning	61
5	MODELAGEM E IMPLEMENTAÇÃO	64
5.1	O SISCONFLUX	65
5.1.1	Interfaces com STVD e SYNCROMAX	67
5.1.2	Modulo de Acompanhamento e Previsão de Cenário - MAPC	67
5.1.3	Módulo de Balanceamento de Fluxo (MBF)	68
5.2	Modelagem do Problema	69
5.2.1	Definição e Características do Agente	70
5.2.2	Natureza do Ambiente	70
5.2.3	Conjunto de Ações	71
5.2.4	Conjunto de Estados	73
5.3	Concepção e Arquitetura do MAAD	75
5.4	Função de Avaliação de Cenário	76
5.4.1	Fator Montante de Atraso - $D(e)$	77
5.4.2	Fator Quantidade de Aeronaves nos Setores ATC - $C(e)$	77
5.4.3	Fator Densidade Dinâmica	78
5.4.4	Avaliação de Cenário	79
5.4.5	Avaliação de Cenário - Pior Caso	81
5.5	Nova Estrutura de Aprendizagem nas Aplicações ATFM	82
6	ESTUDO DE CASO	85
6.1	Planejamento do Estudo de Caso	85
6.2	Configuração dos Cenários	87
6.2.1	Setorização do Espaço Aéreo	87
6.2.2	Fator FPL	88
6.2.3	Cenário de Teste	89
6.2.4	Cenário de Verificação de Convergência	91
6.2.5	Cenário de Aplicação	92
6.2.6	Parâmetros Iniciais de Aprendizagem	93
6.3	Teste Inicial de Validação dos Protótipos	94

6.4	Experimentos e Resultados - Agente ATFM sem a Experiência Humana	97
6.4.1	Verificação de Convergência - Protótipo MAAD	97
6.4.2	Aplicação do Protótipo MAAD	99
6.5	Experimentos e Resultados - Agente ATFM com a Experiência Humana	101
6.5.1	Verificação de Convergência - Protótipo MAAD*	101
6.5.2	Aplicação do Protótipo MAAD*	103
6.6	Discussão de Resultados	105
6.6.1	Tempo de Execução	105
6.6.2	Análise da Convergência	105
6.6.3	Alteração dos Parâmetros de Aprendizagem	108
6.6.4	Comentários Finais	110
7	CONCLUSÃO	112
7.1	Contribuições e Pesquisas Futuras	113
7.2	Considerações Finais	114

Lista de Figuras

1.1.1 Movimento Aéreos	12
1.1.2 Movimentos Aéreos - Crescimento Percentual	13
2.1.1 Quantidade de Aeronaves e Complexidade ATC na FIR CW	23
2.2.1 Subdivisões do Espaço Aéreo em uma FIR (Souza <i>et al</i> , 2008)	24
2.2.2 Regiões de Informação de Vôo e Respectivos Setores de Controle do Brasil	25
2.3.1 Sala de Operações do CGNA	26
2.3.2 Relatório Disponibilizado pelo SYNCROMAX	28
2.4.1 Processo Decisório ATFM (tático) no Brasil	30
2.4.2 Tela do SYNCROMAX Empregada no ATFM tático	31
2.4.3 Tela do STVD	31
3.3.1 Grafo Origem - Destino	38
3.6.1 Sistema Totalmente Automatizado (Agogino e Tumer, 2009)	44
3.6.2 Sistema Semi-automatizado	45
3.6.3 Sistema Semi-automatizado (Agogino e Tumer, 2009)	45
4.4.1 Modelo Padrão de Aprendizagem por Reforço (Kaelbling <i>et al</i> ,1996)	58
5.1.1 Processo Decisório ATFM com o SISCONFLUX	66
5.1.2 Integração do SISCONFLUX com o STVD e SYNCROMAX	66
5.1.3 Arquitetura do MAPC (Souza <i>et al</i> , 2008)	68
5.1.4 Arquitetura do MBF (Souza <i>et al</i> , 2008)	68
5.3.1 Arquitetura do MAAD	75
5.5.1 Sistema Semi-automatizado (Agogino e Tumer, 2009)	83
5.5.2 Proposta do MAAD	84
6.4.1 Convergência - Protótipo MAAD	99
6.4.2 Valores de $Q(s,a)$ - MAAD	101
6.5.1 Convergência - Protótipo MAAD*	102
6.5.2 Valores de $Q(s,a)$ - MAAD*	104
6.6.1 Convergência - Gráfico Comparativo	106
6.6.2 Comportamento - MAAD x MAAD*	107
6.6.3 MAAD - $Q(s,a)$	109

6.6.4 MAAD* - $Q(s,a)$ 109

Lista de Tabelas

5.2.1	Medidas de Ajuste de Fluxo	73
5.2.2	Cenário Aéreo Projetado para a FIR BS em 12/03/2010	74
5.2.3	Cenário Aéreo Projetado para a FIR BS em 12/03/2010 - Com- pilação Simplificada	74
5.4.1	Cálculo de $T(e)$ - Cenário Típico	80
5.4.2	$C(e)$ inerente à FIR BS - 10:15Hs	80
5.4.3	$C(e)$ inerente à FIR CW - 10:15Hs	80
5.4.4	Cálculo de $T(e)$ - Pior Cenário	81
6.2.1	Previsão de Demanda RPL com Fator FPL - FIR BS / FIR CW (06/03/2008)	90
6.2.2	Previsão de Demanda RPL com Fator FPL - FIR RE / FIR AZ (06/03/2008)	90
6.2.3	Previsão de Demanda RPL com fator FPL - FIR BS e FIR CW (18/03/2009)	91
6.2.4	Previsão de Demanda RPL com fator FPL - FIR RE e FIR AZ (18/03/2009)	91
6.2.5	Previsão de Demanda RPL com fator FPL - FIR BS e FIR CW (17/03/2010)	92
6.2.6	Previsão de Demanda RPL com fator FPL - FIR RE e FIR AZ (17/03/2010)	92
6.2.7	Parâmetros de Aprendizagem	94
6.3.1	Medidas de Controle de Fluxo - 1 experimento	95
6.3.2	Medidas de Controle de Fluxo - 1 experimento (continuação)	95
6.3.3	Medidas de Controle de Fluxo - 100 experimentos	96
6.3.4	Medidas de Controle de Fluxo - 100 experimentos (continuação)	96
6.4.1	MAAD - Medidas de Controle de Fluxo / Número de Iterações	98
6.4.2	MAAD - Medidas de Controle de Fluxo - Aplicação	100
6.4.3	MAAD - Medidas de Controle de Fluxo - Aplicação (continuação)	100
6.5.1	MAAD* - Medidas de Controle de Fluxo / Número de Iterações	102
6.5.2	MAAD* - Medidas de Controle de Fluxo - Aplicação	103
6.5.3	MAAD* - Medidas de Controle de Fluxo - Aplicação	104
6.6.1	Tempo de Aprendizagem	105

6.6.2 Parâmetros de Aprendizagem	108
--	-----

Abreviaturas

AI	– Inteligência Artificial (Artificial Intelligence)
ACC	– Centro de Controle de Área (Area Control Center)
ACC-BS	– Centro de Controle de Área de Brasília
ANAC	– Agência Nacional de Aviação Civil
APP	– Controle de Aproximação (Approach Control)
AR	– Aprendizagem por Reforço
ASM	– Gerenciamento do Espaço Aéreo (Air Space Management)
ATC	– Controle de Tráfego Aéreo (Air Traffic Control)
ATFM	– Gerenciamento de Fluxo de Tráfego Aéreo (Air Traffic Flow Management)
ATFMU	– Unidade Central de Gerenciamento de Fluxo de Tráfego Aéreo (Air Traffic Flow Management Central Unit)
ATM	– Gerenciamento de Tráfego Aéreo (Air Traffic Management)
ATPC	– Atraso Total de Peso Combinado
ATS	– Serviços de Tráfego Aéreo (Air Traffic Service)
ATZ	– Zona de Tráfego de Aeródromo (Aerodrome Traffic Zone)
AWY	– Aerovias (Airways)
CATFM	– ATFM Colaborativo (Collaborative Air Traffic Flow Management)
CGNA	– Centro de Gerenciamento da Navegação Aérea
CTA	– Área de Controle (Control Area)
CTR	– Zona de Controle (Control Zone)

DCB	– Balanceamento entre Capacidade e Demanda (Demand and Capacity Balancing)
DECEA	– Departamento de Controle do Espaço Aéreo
FAA	– Federal Aviation Administration
FIR	– Região de Informação de Vôo (Flight Information Region)
FMC	– Célula de Gerenciamento de Fluxo (Flow Management Cell)
FPL	– Plano de Vôo Eventual (Filed Plan)
GHP	– Problema de Espera em Solo (Ground Holding Problem)
HCI	– Interação Homem-Computador (Human-Computer Interaction)
ICAO	– Organização de Aviação Civil Internacional (International Civil Aviation Organization)
INFRAERO	– Empresa Brasileira de Infra-Estrutura Aeroportuária
MAAD	– Módulo de Avaliação e Apoio à Decisão
MAPC	– Módulo de Acompanhamento e Previsão de Cenário
MBF	– Módulo de Balanceamento de Fluxo
MDP	– Processos Decisórios de Markov (Markov Decision Process)
MIS	– Interface Especificada de Módulo (Module Interface Specification)
MPI	– Submódulo de Modelagem e Previsão de Impacto
NuCGNA	– Núcleo do Centro de Gerenciamento da Navegação Aérea
OACI	– Organização da Aviação Civil Internacional
RPL	– Plano de Vôo Repetitivo (Repetitive Plan)
SAD	– Sistema de Apoio à Decisão
SAP	– Submódulo de Aprendizagem
SISCEAB	– Sistema de Controle do Espaço Aéreo Brasileiro
SISCONFLUX	– Sistema de Apoio à Decisão Aplicado ao Gerenciamento Tático do Fluxo de Tráfego
SISDACTA	– Sistema de Defesa Aérea e Controle de Tráfego Aéreo
SISSAR	– Sistema de Busca e Salvamento

SMAC	– Submódulo de Avaliação de Cenário
SPV	– Sistema de Proteção ao Vôo
STA	– Submódulo de Tratamento de Ação
STMA	– Sistema de Telecomunicações do Ministério da Aeronáutica
STP	– Sistema de Tratamento de Plano de Vôo
STR	– Sistema de Tratamento Radar
STVD	– Sistema de Tratamento e Visualização de Dados
SYNCROMAX	– Sistema de Gestão de Fluência de Tráfego Aéreo
TMA	– Área de Controle Terminal (Terminal Maneuvering Area)
TWR	– Torre de Controle (Tower)
UDD	– Unidade de Derivação Digital
UTA	– Área Superior de Controle (Upper Control Area)

Capítulo 1

INTRODUÇÃO

1.1 Motivação

No contexto da recente crise verificada no setor de transporte aeroviário brasileiro, entre os anos de 2006 e 2008, diversos foram os fatores geradores e contribuintes, não cabendo, neste momento, a tentativa de elencá-los na sua totalidade. Contudo, um desses fatores pode estar associado à forma empírica como são utilizadas as técnicas previstas para o controle do fluxo de tráfego, o que se evidencia na aplicação intuitiva de medidas restritivas ao fluxo de tráfego aéreo, com o objetivo de se evitar a saturação dos setores de controle, mormente nos cenários onde se verifica um abrupto desbalanceamento na relação capacidade de controle de tráfego x demanda de fluxo de tráfego, ou seja, em cenários caracterizados como de natureza tática.

A aplicação de medidas restritivas ao fluxo de tráfego, nos cenários táticos, é determinada com a finalidade de atender às prescrições regulamentares em vigor, de forma a impedir que os controladores de vôo trabalhem com carga superior ao estabelecido em legislação, e tem severo impacto sobre todas as atividades operacionais e logísticas aeroportuárias. (Crespo *et al*, 2007).

A escolha das técnicas e o dimensionamento da "dose" a ser aplicada, no que tange às medidas de controle de fluxo de tráfego, são estabelecidos de forma empírica, fortemente condicionada à experiência dos gerentes de fluxo que guarnecem as posições operacionais no Centro de Gerenciamento da Navegação Aérea (CGNA), órgão subordinado ao Departamento de Controle do Espaço Aéreo (DECEA) com a incumbência de desempenhar as atividades relacionadas ao gerenciamento do fluxo de tráfego aéreo no Brasil.

Entretanto, o grande número de fatores e variáveis envolvidos neste processo tornam a tomada de decisão, empiricamente fundamentada, apenas par-

cialmente eficaz. Ou seja, existe espaço para a otimização do fluxo de tráfego mediante o incremento da eficiência das medidas restritivas aplicadas, o que se dá através da escolha sistematizada da medida mais adequada, considerando-se os diversos aspectos envolvidos (capacidade dos setores de controle, dimensão dos setores de controle, demanda por aeroporto, performance das aeronaves, necessidade de priorização de aeroportos, necessidade de priorização de rotas, disponibilidade de pessoal, restrições meteorológicas dentre outros) (Crespo *et al*, 2007).

Em adição, a tendência do volume de tráfego aéreo é de crescimento, conforme pode ser observado no gráfico representado pela Figura 1.1.1, o qual foi construído com dados compilados a partir do sítio na Internet da Empresa Brasileira de Infra-estrutura Aeroportuária (INFRAERO, 2009). No gráfico em tela, que mostra o comportamento da quantidade de movimentos contabilizados apenas nos aeroportos administrados pela INFRAERO, tal tendência pode ser verificada entre os anos de 2003 e 2008; em 2003 o movimento foi pouco superior a 1.500.000 movimentos, ao passo que em 2008 superou 2.000.000 de movimentos.

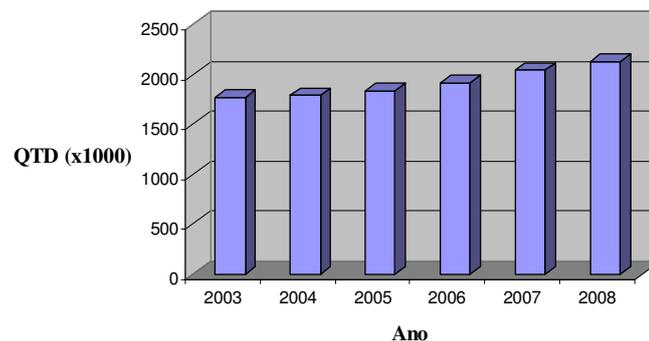


Figura 1.1.1: Movimento Aéreos

Já no gráfico da Figura 1.1.2 , pode-se verificar o comportamento da taxa de crescimento da quantidade de movimentos aéreos contabilizados pela INFRAERO (excluídos os movimentos de aeronaves militares). Entre 2003 e 2008, sempre houve crescimento na quantidade de movimentos aéreos nos aeroportos da INFRAERO. À luz do que foi apresentado, seria desnecessário mencionar a importância atribuída ao desenvolvimento de um sistema que processe esses fatores e variáveis e tenha como saída a medida restritiva de controle de fluxo mais adequada para o cenário vigente.

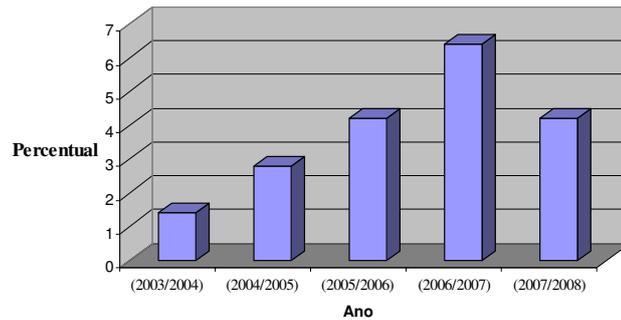


Figura 1.1.2: Movimentos Aéreos - Crescimento Percentual

Sob a perspectiva da Ciência da Computação Aplicada, verifica-se que a Inteligência Artificial tem sido uma preciosa aliada na busca por ferramentas computacionais voltadas para o aprimoramento do Gerenciamento de Tráfego Aéreo (Air Traffic Management – ATM), consideradas todas as suas subdivisões, a saber: Gerenciamento do Espaço Aéreo (Air Space Management – ASM), Controle de Tráfego Aéreo (Air Traffic Control - ATC) e Gerenciamento do Fluxo de Tráfego Aéreo (Air Traffic Flow Management – ATFM).

Segundo a pesquisa realizada, os trabalhos mais recentes discutem a utilização de diversos formalismos, tais como: teoria de autômatos (Bayen *et al*, 2005), agentes inteligentes (Wolf, 2007), aprendizagem por reforço e teoria multi-agentes (Alves, 2006), (Tumer e Agogino, 2007). Então, em 2007 foi proposto o desenvolvimento de um sistema modular para auxiliar os gerentes de fluxo, lotados no CGNA, na aplicação e gerenciamento das medidas de controle de fluxo de tráfego aéreo. Tal sistema, denominado SISCONFLUX (Sistema de Apoio à Decisão Aplicado ao Gerenciamento Tático do Fluxo de Tráfego), é o primeiro fruto de uma proposta estratégica de cooperação entre o Primeiro Centro Integrado de Defesa Aérea e Controle de Tráfego Aéreo (CINDACTA I), a Universidade de Brasília (UnB), o CGNA e a Fundação Atech – Tecnologias Críticas.

1.2 Objetivo

Considerando o problema inerente à aplicação das medidas de controle de fluxo de tráfego aéreo, e no contexto do desenvolvimento do SISCONFLUX, este trabalho tem como objetivo propor e testar um processo de apoio à decisão dotado de capacidade de aprendizagem não supervisionada, de modo que o sistema incorpore a experiência dos usuários, quais sejam, os gerentes de fluxo de tráfego aéreo. Este processo residirá no Módulo de Avaliação e Apoio à Decisão (MAAD) do referido sistema. A implementação deste processo foi

norteada por algumas diretrizes, elencadas a seguir:

- a) Contribuir para a implementação de um sistema de gerenciamento de fluxo de tráfego aéreo que incremente a eficiência das medidas de controle de fluxo de tráfego aplicadas pelo CGNA;
- b) Conceber um processo com capacidade de incorporar a experiência dos especialistas humanos, quais sejam, os gerentes de fluxo de tráfego aéreo do CGNA;
- c) Atribuir ao sistema a capacidade de disponibilizar, para os gerentes de fluxo no CGNA, medidas de controle de fluxo de tráfego geradas segundo critérios pré-estabelecidos, a saber: montante de atraso imputado aos movimentos aéreos afetados pelas medidas e nível de congestionamento dos setores de controle de tráfego aéreo.

Consideradas as diretrizes acima, definem-se a seguir os objetivos específicos do presente trabalho:

- a) Adaptar método de aprendizagem por reforço para modelar o problema ATFM, considerando a definição da arquitetura do Módulo de Avaliação e Apoio à Decisão (MAAD) do sistema em tela;
- b) Desenvolver e testar uma função de avaliação de cenário (estado) que permita calcular as recompensas utilizadas por um algoritmo de aprendizagem por reforço, considerando os critérios (congestionamento de setores de controle e montante de atraso produzido pela medida de controle) definidos pelo CGNA;
- c) Desenvolver e testar um agente inteligente baseado em aprendizagem por reforço concebido para sugerir medidas de controle de fluxo de tráfego aéreo aos gerentes de fluxo do CGNA, e com capacidade para incorporar a experiência dos referidos gerentes;
- d) Estabelecer um estudo comparativo acerca do desempenho do agente considerando duas condições distintas: atuação do agente independentemente da experiência dos gerentes de fluxo e atuação do agente considerando a incorporação da experiência dos gerentes de fluxo.

1.3 Metodologia

O MAAD foi desenvolvido segundo os preceitos teóricos da Inteligência Artificial. Na medida em que o cenário aéreo pode ser modelado como um Processo Decisório de Markov, foi possível desenvolver um agente baseado na Teoria de

Utilidade Multi-atributo, na Teoria da Decisão e na Aprendizagem por Reforço (Russel e Norvig, 2003). Com base nas duas primeiras, foi possível desenvolver uma função de avaliação de estados voltada para a qualificação de cenários ATFM, cujos resultados são utilizados para produzir as recompensas utilizadas pelo agente.

A Aprendizagem por Reforço, por sua vez, possibilitou a implementação de um agente computacional, voltado para o apoio à decisão em um ambiente ATFM, com capacidade para incorporar a experiência dos tomadores de decisão. Desta feita, foi desenvolvido um dispositivo capaz de sugerir políticas de ajuste de fluxo de tráfego aéreo, as quais são disponibilizadas aos gerentes de fluxo (agente humano).

Conforme poderá ser verificado no Capítulo 3, o emprego da Aprendizagem por Reforço nas pesquisas no domínio ATFM é recente. Em (Alves *et al*, 2006, 2008) o formalismo foi utilizado para melhorar a eficiência de um sistema ATFM multi-agente. Neste sistema, a produção de políticas de ajuste de fluxo é tão mais eficiente quanto maior for o número de mensagens trocadas entre os agentes do sistema. Tal condição, contudo, implica restrições de desempenho em termos de capacidade de processamento e tempo computacional. Neste contexto, algoritmos de Aprendizagem por Reforço foram utilizados para otimizar o fluxo de mensagens.

Nos trabalhos mais recentes produzidos por cientistas da NASA e da *Oregon State University* (Agogino e Tumer, 2008, 2009), a Aprendizagem por Reforço foi aplicada em uma ferramenta computacional ATFM de apoio à decisão baseada em técnicas multi-agentes, desenvolvida a partir de sistemas de controle utilizados na robótica.

A partir deste contexto, justifica-se o desenvolvimento do presente trabalho, na medida em que contribui para a busca de soluções próprias, suportadas pela Inteligência Artificial, para uma questão de grande relevância para o país, mormente quando considerado potencial incremento na demanda causado por eventos como a Copa do Mundo de 2014 e as Olimpíadas de 2016.

1.3.1 Metodologia de Pesquisa e Implementação

Para a realização do trabalho, foram realizadas visitas ao CGNA e encontros freqüentes com engenheiros e analistas da empresa Atech, responsáveis pelo desenvolvimento do Sistema para o Gerenciamento de Fluxo de Tráfego Aéreo - SYNCROMAX (Timoszczuk *et al*, 2009), uma das ferramentas ATFM uti-

lizadas no cenário nacional. O CGNA disponibilizou toda a sua base de dados concernentes ao histórico de demanda por fluxo de tráfego, bem como todos os relatórios referentes às medidas de controle de fluxo de tráfego aplicadas no país no período de janeiro de 2008 a março de 2010. A partir deste material, foi possível criar os cenários de treinamento e de teste do agente. O histórico das medidas de controle de fluxo de tráfego permitiu construir um conjunto de ações (para o agente computacional) que agregou a experiência do gerente de fluxo lotado no CGNA (agente humano).

O desenvolvimento do projeto incluiu as seguintes etapas:

- Identificação e descrição da problemática ATFM, de forma que os objetivos do projeto pudessem ser definidos; nesta fase, foram consultadas as diversas publicações internas ao Comando da Aeronáutica versando sobre Controle de Tráfego Aéreo e Gerenciamento de Fluxo de Tráfego Aéreo; também foram realizadas visitas ao Centro de Gerenciamento da Navegação Aérea, com o objetivo de pesquisar a dinâmica do processo decisório ATFM no país, bem como conhecer as ferramentas que apóiam tal processo;
- Realização de pesquisa bibliográfica referente a pesquisas conduzidas no domínio do Gerenciamento de Tráfego Aéreo (ATM), com ênfase para trabalhos na área do Gerenciamento de Fluxo de Tráfego Aéreo (ATFM); nesta fase foi possível verificar as diversas metodologias e técnicas empregadas na busca por soluções para os problemas ATFM; neste ponto foi possível confirmar a importância da temática ATFM no contexto científico internacional, com uma infinidade de trabalhos publicados segundo linhas de pesquisa desenvolvidas nas engenharias e na computação;
- Definição da solução utilizada para alcançar os objetivos definidos a partir da identificação da problemática ATFM; neste momento optou-se pela utilização da Inteligência Artificial e, de forma mais específica, da Aprendizagem por Reforço na concepção de um agente inteligente capaz de sugerir medidas restritivas de controle de fluxo de tráfego aéreo.
- Definição de uma função de avaliação que contemplasse os critérios utilizados pelos gerentes de fluxo de tráfego no processo de estabelecimento das políticas de controle de fluxo de tráfego; a partir dos conceitos inerentes à Teoria da Utilidade Multi-atributo e da Teoria de Decisão, foi definida uma função para avaliar o cenário aéreo e nortear as ações do agente computacional; tanto quanto possível, esta função reflete os interesses dos gerentes de fluxo de tráfego na busca pela condição de balanceamento entre demanda de fluxo e capacidade do espaço aéreo;

- Definição da arquitetura do protótipo do agente, incluindo a definição do algoritmo a ser empregado na solução do problema de Aprendizagem por Reforço (avaliação da política e definição da política ótima); nesta fase foi apresentada a arquitetura do protótipo incorporando o algoritmo baseado em método de Diferença Temporal denominado Q-learning; a escolha do algoritmo se deu em função dos fatores elencados a seguir: a) obtenção de resultados consistentes na sua aplicação em outras linhas de pesquisa desenvolvidas pelo grupo, tais como Alves (2006) e Arruda (2009); b) diferentemente das soluções envolvendo Programação Dinâmica, não há a necessidade de se conhecer, modelar matematicamente e processar o modelo de transição de estados; c) diferentemente das soluções empregando o Método de Monte Carlo, o Q-learning é um algoritmo *single-step*, e tal característica é fundamental, dada a dinâmica dos cenários ATFM;
- Implementação dos protótipos e realização do estudo de caso; com o objetivo de assegurar a compatibilidade com outros sistemas utilizados pelo CGNA, o protótipo foi desenvolvido em linguagem de programação Java; o protótipo funciona de forma interativa com outros dois módulos desenvolvidos pelo grupo, conforme poderá ser verificado posteriormente; para a realização do estudo de caso, foram utilizados os registros de demanda (planos de voo repetitivos - RPL) referentes aos anos de 2008, 2009 e os três primeiros meses de 2010, a partir dos quais foi possível criar os cenários de treinamento e de aplicação; o conjunto de ações do protótipo autônomo (desprovido da experiência do especialista humano) foi populado pelo Módulo de Balanceamento de Fluxo (MBF); a partir dos Relatórios Diários de Situação, obtidos junto ao CGNA, foi possível criar o conjunto de ações do protótipo que incorpora a experiência dos gerentes de fluxo.

1.4 Organização do Trabalho

O presente trabalho está estruturado em 7 capítulos, através dos quais são apresentadas as motivações e justificativas para a realização do mesmo, as características do cenário ATFM nacional, incluindo a problemática ATFM, e a solução proposta para concluir o desenvolvimento de uma ferramenta computacional ATFM, de apoio à decisão, a ser utilizada no CGNA. A seguir, sintetiza-se o conteúdo dos capítulos seguintes.

Capítulo 2: Neste capítulo é apresentado o *background* necessário à compreensão do problema que este trabalho se propõe a resolver. Neste contexto, são elencados os principais conceitos inerentes ao ATM e discorre-se sobre a

constituição do Sistema de Controle do Espaço Aéreo Brasileiro – SISCEAB. Comenta-se ainda sobre a sistemática do ATFM no Brasil e finalmente é apresentada a problemática ATFM.

Capítulo 3: Nesta parte do trabalho, é apresentada uma revisão dos principais trabalhos publicados tratando do domínio ATFM. São apresentadas algumas linhas de pesquisas que objetivaram o desenvolvimento de ferramentas computacionais aplicadas na solução de problemas ligados ao tráfego aéreo. Neste capítulo estão incluídos também os mais recentes trabalhos escritos e publicados tratando da questão em pauta.

Capítulo 4: Neste capítulo é apresentada a fundamentação teórica concernente à metodologia e aos formalismos de Inteligência Artificial utilizados no desenvolvimento da ferramenta computacional proposta. Ou seja, são apresentados os principais aspectos concernentes à Teoria da Utilidade, à Teoria da Decisão, ao Processo Decisório, à Aprendizagem por Reforço (AR) e ao Algoritmo Q-Learning.

Capítulo 5: Aqui apresenta-se uma visão geral do Sistema de Apoio a Decisão Aplicado ao Gerenciamento Tático do Fluxo de Tráfego, seguida pela concepção e arquitetura do protótipo desenvolvido (MAAD). Neste capítulo encontra-se também a função de avaliação de cenário empregada no processo de aprendizagem concebido para o módulo. Discorre-se ainda sobre a implementação do módulo, tratando-se do ambiente de desenvolvimento e do ambiente de execução. Também são feitos comentários sobre os principais aspectos referentes à integração do MAAD com o SISCONFLUX, tais como formato das bases de dados, taxas de atualização do sistema, consistência da base de dados, etc.

Capítulo 6: Esta é a parte do trabalho que contém o estudo de caso realizado para avaliar o MAAD. Os estudos contemplam o processamento de cenários reais gerados a partir da base de dados obtida junto ao CGNA. Esta base abarca grupos de vôos realizados pelas linhas aéreas comerciais no período compreendido entre janeiro de 2008 e março de 2010.

Capítulo 7: Na parte final do trabalho, é apresentado um resumo das contribuições da proposta, incluindo ainda a proposição de pesquisas futuras.

Capítulo 2

GERENCIAMENTO DO TRÁFEGO AÉREO

2.1 Conceitos Fundamentais

O Gerenciamento de Tráfego Aéreo (ATM - *Air Traffic Management*) tem por objetivo permitir que os exploradores das aeronaves possam cumprir com os horários previstos de saída / chegada e manter os perfis ótimos de voo com o mínimo de limitações sem, contudo, afetar os padrões de segurança. Três elementos principais fazem parte do ATM: Controle de Tráfego Aéreo (ATC - *Air Traffic Control*), Gerenciamento do Espaço Aéreo (ASM - *Air Space Management*) e Gerenciamento de Fluxo de Tráfego Aéreo (ATFM - *Air Traffic Flow Management*), garantindo assim o movimento seguro, ordenado, rápido, compatível com as necessidades ambientais e econômicas da aeronave durante todas as fases do voo.

Os principais objetivos de cada um dos elementos que constituem o ATM são especificados abaixo:

ASM: prover, considerada a estrutura do espaço aéreo (setores de controle, aerovias, áreas terminais, etc), um nível de disponibilidade suficiente para atender a demanda de tráfego;

ATC: separar as aeronaves para evitar colisões, organizar e agilizar o fluxo de tráfego e fornecer informações e outros apoios para os pilotos, quando pertinente;

ATFM: garantir o melhor fluxo de tráfego possível em situações onde a demanda excede a capacidade disponível do sistema ATC.

A presente proposta foi concebida no contexto ATFM, posto que as medidas de controle de fluxo de tráfego constituem ações aplicáveis em uma condição

de desbalanceamento entre a demanda de fluxo de tráfego e a capacidade ATC disponível. Conforme definido pela Organização da Aviação Civil Internacional-OACI (ICAO, 2005), o ATFM, mais recentemente definido como DCB (*Demand and Capacity Balancing*) é aplicado segundo três fases, especificadas a seguir.

- Estágio Estratégico: a ação é realizada mais de um dia antes do dia em que terá efeito; o planejamento estratégico é normalmente realizado com bastante antecedência, normalmente de dois a seis meses antes.
- Planejamento Pré-tático: a ação deve ser tomada no dia anterior ao dia em que entrará em vigor.
- Operações Táticas: a ação deve ser tomada no dia em que entrará em vigor.

2.1.1 Planejamento Estratégico

O planejamento estratégico deve ser realizado em conjunto com os órgãos ATC e os operadores de aeronaves. Consiste em analisar a demanda para um determinado período futuro, avaliar onde e quando a demanda pode superar a capacidade ATC disponível e, se necessário, tomar medidas para resolver o desequilíbrio. Algumas dessas medidas são apresentadas abaixo.

- Verificar e coordenar, junto a autoridade ATC, a possibilidade de prover a capacidade adequada no local e na hora prevista.
- Reordenar determinados fluxos de tráfego (criação de novas rotas).
- Reordenar os horários dos vôos, conforme a necessidade.
- Identificar a necessidade de se adotar medidas restritivas para controlar fluxo de tráfego.

2.1.2 Planejamento Pré-tático

O Planejamento Pré-tático deve implicar o ajuste fino do plano estratégico, com base em informações atualizadas sobre a demanda. Medidas típicas desta fase são descritas a seguir.

- Reordenar certos fluxos de tráfego (criação de novas rotas)
- Reordenar certos fluxos de tráfego (rotas alternativas).
- Prever a adoção de medidas restritivas para controlar o fluxo de tráfego.
- Publicar, com a antecedência necessária, as medidas planejadas.

2.1.3 Operações Táticas

As ações ATFM táticas previstas são elencadas abaixo.

- Executar as medidas ATFM táticas previstas, com o propósito de regular o fluxo de tráfego e não exceder a capacidade ATC disponível.
- Acompanhar a evolução do cenário aéreo, objetivando a certificação de que as medidas ATFM aplicadas estão tendo os efeitos desejados e, caso necessário, adotar as ações corretivas.

2.1.4 Capacidade ATC

Conforme definido pelo Departamento de Controle do Espaço Aéreo (DECEA, 2007), a Capacidade ATC representa a medida de habilidade do órgão ATC ou de suas posições operacionais em prover serviço, em condições normais, para as aeronaves. A capacidade ATC é expressa como o número de aeronaves operando simultaneamente em determinada porção de espaço aéreo, levando em conta as condições meteorológicas, a configuração do órgão ATC, o pessoal e os equipamentos disponíveis, bem como quaisquer outros fatores que possam afetar a carga de trabalho do controlador de tráfego aéreo responsável pelo espaço aéreo. Ou seja, por capacidade ATC entende-se a quantidade de movimentos aéreos que um setor de controle de tráfego aéreo comporta de forma simultânea. Tal capacidade se define a partir de características específicas, a saber:

- Tipo de elemento regulado – diz respeito ao tipo de setor e ou área na qual o serviço de ATC é prestado; pode ser uma Área Terminal de Aproximação (TMA - *Terminal Maneuvering Area*) ou um setor de região de informação de voo.
- Quantidade de coordenações – diz respeito à quantidade de comunicações necessárias para prestar o serviço de controle de tráfego aéreo; estas comunicações podem ser entre controladores e aeronaves e entre os controladores de diferentes centros ATC.
- Pontos de conflito – esta característica se define pelos pontos onde ocorrem cruzamentos de trajetória das aeronaves, considerando a circulação aérea interior ao elemento regulado; estes pontos de cruzamento poder ser verificados nos planos horizontal e vertical.
- Tempo de permanência – esta característica leva em consideração as dimensões do setor de controle, o que determina o tempo que os movimentos aéreos levam trafegando no interior do mesmo.

Como consequência, infere-se que os valores das capacidades ATC poderão ser alterados em função de eventuais degradações, inoperâncias ou indisponibilidades dos elementos que constituem a infra-estrutura do Sistema de Controle do Espaço Aéreo – SISCEAB. A Capacidade ATC, contudo, está fortemente correlacionada com um outro conceito de fundamental importância para o desempenho tanto dos serviços de controle de tráfego aéreo quanto do serviço de gerenciamento de fluxo de tráfego, qual seja, a Complexidade do Espaço Aéreo.

2.1.5 Complexidade do Espaço Aéreo (Complexidade ATC ou Densidade Dinâmica)

Segundo a Federal Aviation Administration (FAA, 2001), a Complexidade do Espaço Aéreo pode ser definida como o efeito coletivo dos fatores e/ou variáveis (fatores de complexidade) que contribuem para tornar a atividade de controle de tráfego aéreo, em uma determinada porção do espaço aéreo e em um determinado período, mais difícil para os controladores. O grau de dificuldade imputado à atividade ATC está comumente associado a fatores estruturais fixos, inerentes ao setor de controle, e a fatores variáveis, inerentes ao fluxo de aeronaves, conforme pode ser observado a seguir (Sridhar *et al*, 1998).

- Características estruturais do espaço aéreo (dimensões do setor de controle, quantidade e disposição das aerovias, auxílios a navegação disponíveis).
- Características do fluxo de tráfego (quantidade de aeronaves, *mix* de aeronaves, proximidade de terminais de controle, tipo de serviço ATC prestado e separações aplicadas, condições meteorológicas, medidas de controle de fluxo em vigor).

A complexidade ATC é crucial porque tem implicação direta sobre a carga de trabalho dos controladores de tráfego aéreo, ou seja, constitui critério crítico para a segurança no controle do espaço aéreo. A Figura 2.1.1 apresenta o comportamento da quantidade de aeronaves e da complexidade ATC de um setor de controle da Região de Informação de Vôo de Curitiba (FIR CW), referente ao dia 07 de julho de 2009.

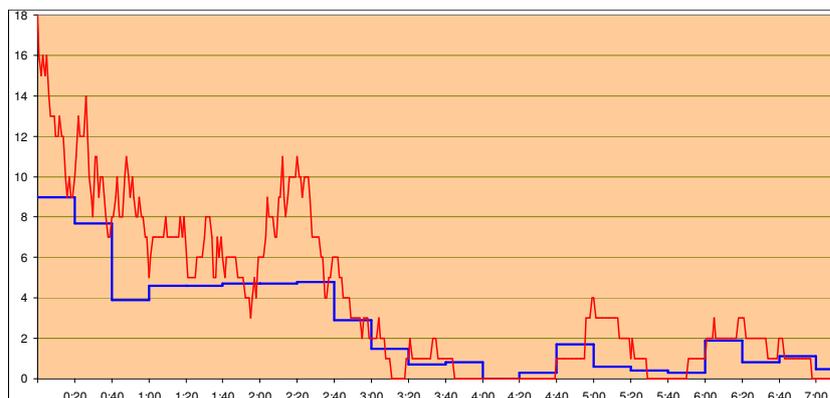


Figura 2.1.1: Quantidade de Aeronaves e Complexidade ATC na FIR CW

Na figura 2.1.1, a linha vermelha representa a quantidade de aeronaves no setor, em função da hora do dia. A linha azul apresenta um valor relativo referente à Complexidade ATC. Por exemplo, no horário compreendido entre 05:00 e 05:20, a quantidade de aeronaves no setor é superior à quantidade observada no período de 06:00 e 06:20; contudo, neste último período, a complexidade é maior. Tal condição é fundamental no dimensionamento e aplicação de medidas de controle de fluxo, pois tem impacto direto na carga de trabalho do controlador de tráfego aéreo, na medida em que maior complexidade, via de regra, implica maior carga de trabalho.

2.2 O SISCEAB

Em 1990, foi instituído o Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB), tendo por finalidade integrar os seguintes Sistemas: de Proteção ao Voo (SPV), de Telecomunicações do Ministério da Aeronáutica (STMA), de Busca e Salvamento (SISSAR) e de Defesa Aérea e Controle de Tráfego Aéreo (SISDACTA). O SISCEAB é gerenciado pelo Comando da Aeronáutica, através do órgão central do Sistema, o Departamento de Controle do Espaço Aéreo-DECEA (Crespo *et al*, 2007).

O DECEA gerencia o espaço aéreo do Brasil através de seus órgãos Regionais, destacando-se os CINDACTA (Centros Integrados de Defesa Aérea e Controle do Espaço Aéreo), e os permanentes do SISDACTA que têm a incumbência de manter seguro e confiável o fluxo do tráfego aéreo nas áreas sob sua jurisdição. Os CINDACTA têm como atribuição principal o desenvolvimento da atividade de Controle de Tráfego Aéreo (ATC), contudo, contando com auxílio de Células de Gerenciamento de Fluxo (FMC) instaladas localmente, auxiliam o CGNA no Gerenciamento do Fluxo de Tráfego (ATFM) pré-tático e atuam no ATFM tático.

2.2.1 Estrutura do Espaço Aéreo

Com a finalidade de possibilitar a organização, o controle e o gerenciamento do tráfego aéreo, o espaço aéreo é dotado de uma estrutura própria, composta pelas aerovias e demais subdivisões, tais como: Região de Informação de Vôo (FIR), Setores de Controle, Área de Controle Terminal (TMA), Zona de Controle (CTR), etc. As aerovias podem ser definidas como caminhos (ou estradas) tridimensionais dentro dos quais as aeronaves devem trafegar, ao passo que as demais subdivisões do espaço aéreo foram criadas com o intuito de delimitar a área de jurisdição dos diversos órgãos de controle, assim como definir regiões onde determinadas regras operacionais de controle têm vigência. Figura 2.2.1 mostra, de forma simplificada, como se subdivide o espaço aéreo delineado por uma FIR. Ainda tomando como referência a Figura 2.2.1, destaca-se que os grandes aeroportos são envolvidos pelas TMA/CTR, sob responsabilidade de um Centro de Controle de Aproximação (APP). No Brasil existem 39 Áreas de Controle Terminal.

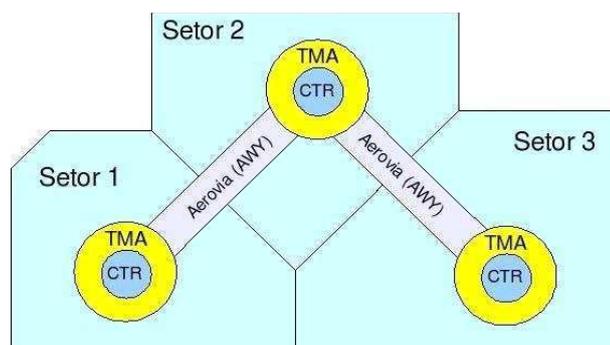


Figura 2.2.1: Subdivisões do Espaço Aéreo em uma FIR (Souza *et al.*, 2008)

O espaço aéreo brasileiro é dividido em cinco Regiões de Informação de Vôo (FIR), sob responsabilidade de quatro CINDACTA. Conforme comentado previamente, as Regiões de Informação de Vôo são jurisdicionadas a um determinado CINDACTA, ou seja: a FIR Brasília (FIR BS / SBBS) ao CINDACTA I, a FIR Curitiba (FIR CW / SBCW) ao CINDACTA II, a FIR Recife (FIR RE / SBRE) e a FIR Atlântico (FIR AO / SBAO) ao CINDACTA III e a FIR Amazônica (FIR AZ / SBZ) ao CINDACTA IV. A Figura 2.2.2 mostra a divisão do espaço aéreo brasileiro em Regiões de Informação de Vôo e em setores de controle.

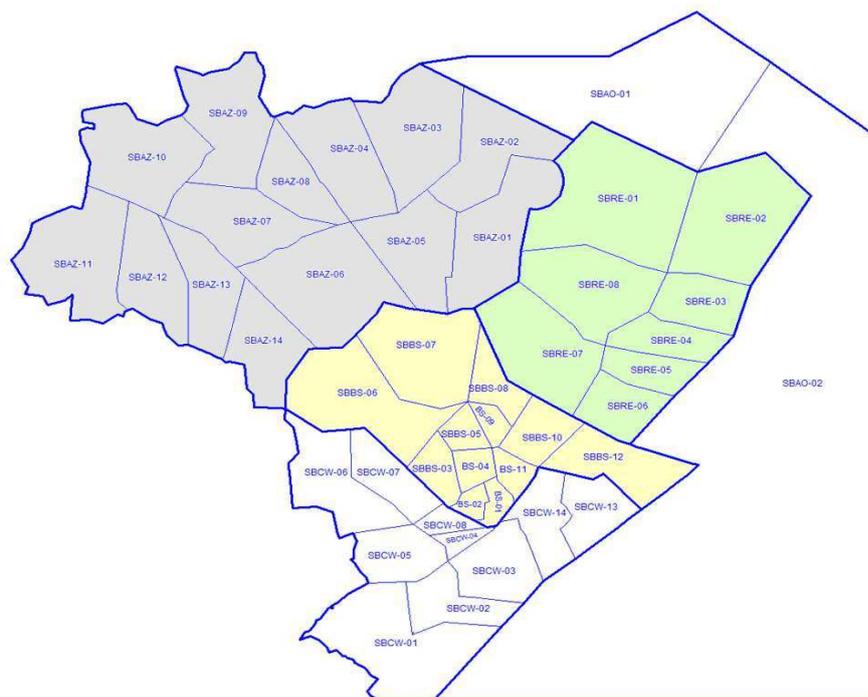


Figura 2.2.2: Regiões de Informação de Voo e Respectivos Setores de Controle do Brasil

As FIR são subdivididas em setores de controle. Via de regra, cada setor de controle, incluindo as aerovias, fica sob responsabilidade de dois controladores e um supervisor, lotados nos Centros de Controle de Área (ACC). Cada CINDACTA possui um ACC com responsabilidade sobre a FIR correspondente. O espaço aéreo brasileiro está dividido em 46 setores de controle (14 na FIR Amazônica, 12 na FIR Brasília, 10 na FIR Curitiba, 8 na FIR Recife e 2 na FIR Atlântico).

2.3 ATFM no Brasil

No âmbito do SISCEAB, o início das atividades ATFM remonta ao ano de 1996, quando surgiu a necessidade de se coordenar o aeroporto de Congonhas. Denomina-se como “coordenado” o aeroporto cuja demanda de tráfego aéreo é superior à capacidade adotada, tendo, portanto, todas as suas operações de pouso e decolagem condicionadas à obtenção de *slot* ATC. Ou seja, todos os pousos e decolagens são restritos a horários pré-estabelecidos e previamente alocados a todos os exploradores (empresas aéreas, empresas de taxi-aéreo) e demais usuários. Em 1995, foi criado o Núcleo do Centro de Gerenciamento da Navegação Aérea (NuCGNA), em São José dos Campos, com a finalidade de assessorar o DECEA nos assuntos pertinentes ao ATFM. Em 2006, o Núcleo foi transferido para o Rio de Janeiro e, em 31 de agosto de 2007 foi ativado o CGNA, com a incumbência de atuar como órgão executivo e com a estrutura

e atribuições verificadas na atualidade. A Figura 2.3.1 mostra a sala de operações do CGNA, no Rio de Janeiro.



Figura 2.3.1: Sala de Operações do CGNA

Conforme mencionado acima, no Brasil, o CGNA tem a incumbência de desempenhar as atividades ATFM em âmbito nacional. E para fazer frente ao ATFM, nos níveis estratégico, pré-tático e tático, o CGNA conta hoje com uma estrutura que agrega uma Unidade Central de Gerenciamento de Fluxo (ATFMU), denominada Seção de Gerenciamento de Fluxo de Tráfego Aéreo, e Células de Gerenciamento de Fluxo (FMC) localizadas nos CINDACTA (Miyamaru, 2008). A ATFMU e as FMC utilizam as informações produzidas pela ferramenta SYNCROMAX (Timoszczuk *et al*, 2009), de forma que cenários possam ser projetados e as ações devidas sejam tomadas tempestivamente, evitando o desbalanceamento entre demanda e capacidade de fluxo.

Dentre as principais responsabilidades do CGNA, pode-se destacar (DECEA, 2007):

- a) conceber e aplicar programas nacionais de gerenciamento de fluxo de tráfego aéreo;
- b) monitorar e analisar os componentes do SISCEAB e as condições meteorológicas, verificando impactos potenciais sobre o fluxo de tráfego;
- c) ser o ponto focal para a regulação diária das funções ATFM;
- d) determinar quando a capacidade do SISCEAB será impactada a ponto de demandar a implementação de um procedimento de gerenciamento de fluxo de tráfego aéreo;
- e) implementar procedimentos nacionais de gerenciamento de fluxo de tráfego aéreo, quando necessário, para ordenar o fluxo do tráfego aéreo no SISCEAB;

- f) recomendar e aprovar alternativas de gerenciamento de fluxo de tráfego aéreo, quando procedimentos nacionais não forem adequados;
- g) monitorar a efetividade dos procedimentos de gerenciamento de fluxo de tráfego aéreo aplicados no SISCEAB, modificando-os ou cancelando-os quando necessário;
- h) avaliar a propriedade da aplicação dos procedimentos de gerenciamento de fluxo de tráfego aéreo entre os órgãos de controle, tendo ascendência operacional para arbitrar o procedimento mais adequado;
- i) definir, por um período de tempo especificado, os valores das capacidades praticadas de aeroportos ou de setores ATC, em função da disponibilidade de elementos de suas infra-estruturas instaladas;
- j) emitir a documentação relativa à declaração de aeroportos coordenados e de aeroportos monitorados;
- l) alocar os *SLOTS* ATC para os aeroportos coordenados e/ou monitorados;
- m) divulgar os *SLOTS* ATC alocados relativos aos aeroportos coordenados para as Salas AIS e TWR daqueles aeroportos; e
- n) controlar o uso dos *SLOTS* ATC alocados nos aeroportos coordenados e monitorados.

Na fase estratégica, as medidas ATFM redundam em ações de planejamento (médio e longo prazo). Neste contexto, as demandas das empresas aéreas (transporte aéreo regular) são recebidas e analisadas pelo CGNA e, a partir de projeções feitas através do SYNCROMAX, levando-se em consideração ainda fatores de correção de demanda determinados pela análise estatística dos demais vôos (aviação geral e militar), as solicitações das empresas são acatadas (integralmente ou com ajustes) ou rejeitadas, conforme a capacidade da infraestrutura disponível (capacidade ATC e capacidade de pista) e, finalmente, a malha de fluxo aéreo é dimensionada para determinado período futuro.

Na fase pré-tática, ainda empregando a ferramenta SYNCROMAX, o CGNA, em coordenação com as empresas aéreas e com os órgãos ATC, projeta medidas conjunturais e pontuais voltadas para o ajuste de condições de desbalanceamento capacidade x demanda previsíveis (e que ainda não vigoravam na fase de planejamento estratégico), tais como manutenções (previstas em calendário) dos equipamentos de suporte ao ATC (manutenções preventivas) e

manutenções corretivas não críticas. A Figura 2.3.2 apresenta a forma de apresentação de dados disponibilizada pelo SYNCROMAX usualmente utilizada. Na referida figura, que apresenta as informações referentes a um setor de controle, o eixo horizontal registra o horário e no eixo vertical observa-se a quantidade de aeronaves no setor (nos diversos horários). Pode-se notar, por exemplo, que às 17:15Hs ocorreu um pico de 13 movimentos no setor de controle.

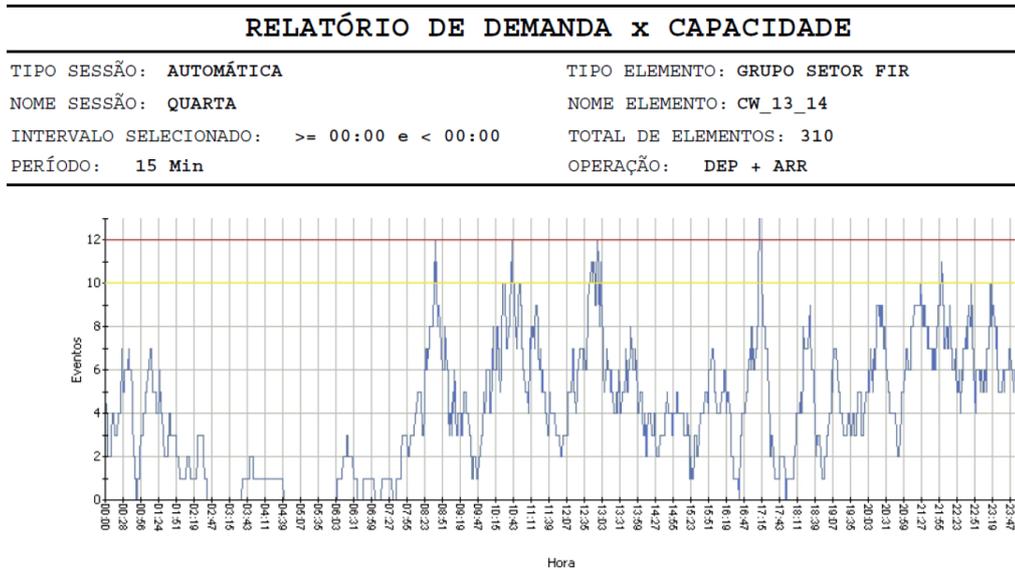


Figura 2.3.2: Relatório Disponibilizado pelo SYNCROMAX

Uma vez planejada e aprovada, a malha de fluxo então é voada, e os órgãos de controle de tráfego, nas respectivas áreas de jurisdição e através dos controladores de vôo, tem a responsabilidade de controlar e ordenar os movimentos aéreos, provendo as separações e as informações necessárias para que o vôo transcorra com segurança. Neste contexto, os controladores de vôo mantêm um fluxo de tráfego que foi planejado levando-se em consideração fatores (estrategicamente dimensionados) com razoável grau de previsibilidade, tais como: disponibilidade de meios técnico-operacionais (radares, frequências de comunicação, consoles operacionais para visualização, etc.), dimensões dos setores de controle, complexidade dos setores de controle, e que pode ser (taticamente) impactado por outros fatores, com menor grau de previsibilidade, tais como condições meteorológicas e experiência profissional do pessoal ATC.

2.4 Problemática ATFM nas Operações Táticas

Em condições normais, onde não se verificam problemas, no suporte a operação ATC, com implicação sobre a infra-estrutura disponível (capacidade ATC e de pista), a demanda de fluxo é atendida sem que sejam gerados atrasos nos horários de decolagem previstos. Contudo, problemas relacionados aos fatores

supracitados podem gerar um desbalanceamento entre a demanda de fluxo e a capacidade de controle, redundando na saturação de determinado(s) setor(es) de controle, o que implica a necessária adoção de medidas restritivas de fluxo de tráfego, ou seja, medidas ATFM no nível tático.

Neste contexto, cabe ressaltar que, segundo as normas brasileiras em vigor, a saturação dos setores de controle se caracteriza pela permanência simultânea de 14 ou mais aeronaves em um setor desagrupado, e de 12 aeronaves em um setor agrupado (setor formado pelo agrupamento de dois ou mais setores). Contudo, uma flexibilização dos parâmetros de saturação de um setor pode ser determinada por diversos fatores, dentre os quais destacam-se: as dimensões do setor, a posição geográfica, quantidade de aerovias, cruzamentos de aerovias no setor, etc.

Conforme estabelecido pelo Departamento de Controle do Espaço Aéreo (DECEA, 2007), é de responsabilidade do CGNA, através de monitoramento constante e contando com o suporte das Células de Gerenciamento de Fluxo locais (FMC), a identificação das condições e momentos em que há risco de saturação dos setores de controle e, a partir daí, adotar medidas restritivas ao fluxo de tráfego, a fim de evitar que os limites de segurança sejam extrapolados. A legislação em vigor prevê a aplicação das seguintes medidas:

- esperas no solo (*ground holding delay*)
- esperas em rota (*airborne holding delay*)
- redução de velocidade
- rotas alternativas
- vetoração para atrasos em rota
- pouso e espera em aeródromos intermediário.

Com referência às aplicações constantes no presente trabalho, apenas a medida espera em solo será empregada, e a mesma se define pela aplicação de atrasos nos horários de decolagem de aeronaves decolando do(s) aeroporto(s) contido(s) em uma TMA, ou grupo de TMA.

Para a aplicação das medidas ATFM de ordem tática, o CGNA utiliza as informações disponibilizadas pelo Sistema de Tratamento e Visualização de Dados (STVD) e pelo SYNCROMAX. O STVD é uma ferramenta computacional ATC desenvolvida para ser utilizada nos centros de controle de tráfego aéreo. Desta feita, o STVD permite a visualização dos dados relativos aos planos de

vôos pré-ativos no sistema (vôos prestes a decolar) e a visualização da síntese radar (aeronaves em vôo).

O SYNCROMAX permite a visualização da demanda prevista nos períodos posteriores ao momento em que ocorreu a degradação da infra-estrutura. Contudo, nenhuma das duas ferramentas tem a capacidade de avaliar o cenário aéreo e, em função das suas características, sugerir medidas de controle de fluxo dimensionadas para evitar que os setores de controle congestionem ou saturem. Esta avaliação é feita por um agente humano, o Gerente de Fluxo.

O processo decisório ATFM tático pode ser melhor compreendido através da Figura 2.4.1. Na referida figura, verifica-se que o Gerente de Fluxo atua no processo decisório apoiado por procedimentos padronizados sugeridos para determinadas configurações do cenário (problema identificado), de acordo com as informações obtidas por meio do SYNCROMAX e do STVD. Tais procedimentos padronizados resultam de experiências prévias catalogadas pelos próprios Gerentes de Fluxo. Obviamente, não é possível a um agente humano avaliar, com eficácia e eficiência, o complexo cenário aéreo brasileiro, de modo que as decisões nesse processo certamente podem ser otimizadas.

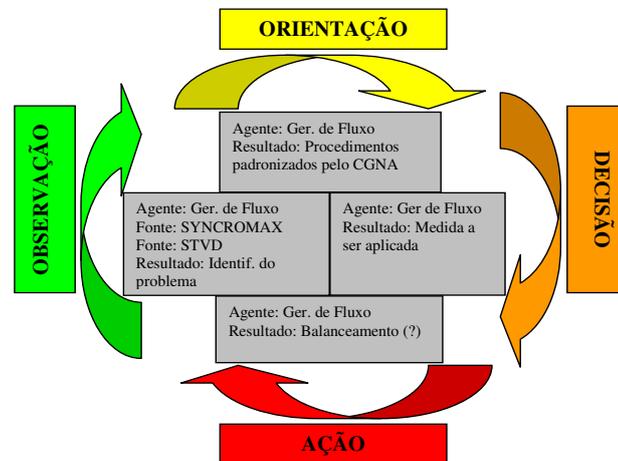


Figura 2.4.1: Processo Decisório ATFM (tático) no Brasil

As Figuras 2.4.2 e 2.4.3 apresentam as telas do SYNCROMAX e do STVD, respectivamente, utilizadas pelo Gerente de Fluxo para realizar o ATFM tático.

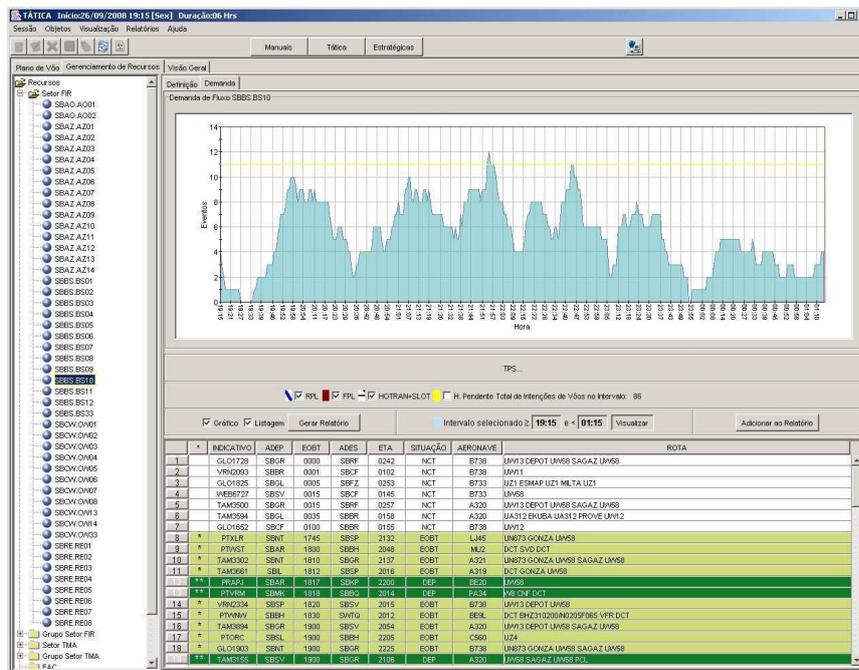


Figura 2.4.2: Tela do SYNCROMAX Empregada no ATFM tático

Na figura 2.4.2, são apresentadas informações acerca da relação capacidade x demanda (hora a hora). Também são apresentadas as intenções de voo (RPL e FPL) e a confirmação de que os vôos já foram autorizados a decolar pelas Torres de Controle. No *frame* a esquerda da tela, pode-se selecionar o setor (ou agrupamento de setores) analisado. Na Figura 2.4.3, são observados os movimentos aéreos nos setores de controle.



Figura 2.4.3: Tela do STVD

Em suma, o problema geral apresentado neste trabalho diz respeito ao fato de que o CGNA não dispõe de um sistema específico voltado para o gerenciamento tático do fluxo de tráfego. E a necessidade de tal sistema é mais evidente na vigência de cenários onde o desbalanceamento ocorre de forma abrupta, decorrente de eventos críticos tais como acidentes e grandes tempestades. Assim sendo, acrescenta-se que o SISCONFLUX foi concebido para atuar nesta

lacuna, vindo a constituir uma ferramenta de apoio à decisão automatizada para ser empregada de forma semi-automatizada, pois, em última instância, a decisão sobre a aplicação da medida de controle processada e sugerida recairá sobre o agente humano.

O problema do ATFM tático pode ser analisado sob a forma de problemas específicos, relacionados com as possíveis medidas de controle aplicáveis (cada tipo de medida corresponde a um problema específico). E tais problemas específicos, sem exceção, podem ser subdivididos em três subproblemas, a saber: compilação e análise do cenário aéreo (identificação do desbalanceamento), dimensionamento da medida aplicada (quanto tempo de espera em vôo, quanto tempo de espera no solo, qual rota alternativa, etc), e finalmente a análise do impacto da medida, decorrendo a avaliação da adequabilidade da medida segundo critérios específicos, definidos pelo usuário.

Capítulo 3

TÉCNICAS COMPUTACIONAIS APLICADAS AO ATFM

O gerenciamento de fluxo de tráfego aéreo (Air Traffic Flow Management – ATFM) é um problema que envolve sincronização em tempo real (Stoltz e Ky, 2001, Stoltz e Guerreau, 2002). Muitas pesquisas vem sendo realizadas com o intuito de resolver este problema, porém, como todo problema crítico, as soluções computacionais esbarram na eficiência dos sistemas computacionais, na necessidade de escalabilidade, de alta segurança e sincronização. Neste contexto, a busca por soluções computacionais para problemas no domínio ATFM representa um importante desafio para a comunidade científica.

A pesquisa bibliográfica realizada até o momento sinaliza que as técnicas mais utilizadas na tentativa de resolução dos problemas ATFM podem agregar Programação Linear Inteira (Rizzi, 2003; Ball *et al*, 2003; Mukherjee, 2004), Programação Dinâmica (Dell’Olmo e Lulli, 2003; Zhang *et al*, 2005) e formalismos de Inteligência Artificial, tais como Sistemas Especialistas (Weigang *et al*, 1997) e, mais recentemente, Aprendizagem por Reforço (Alves *et al*, 2008; Agogino e Tumer, 2009).

Consideradas as propostas no campo da Inteligência Artificial, pode-se afirmar que o ATFM constitui um complexo processo de tomada de decisão que envolve várias entidades, onde agentes computacionais inteligentes podem ser empregados tanto em simulações quanto em operações reais, apoiando operadores humanos (Wolf, 2007). Neste contexto, agentes computacionais podem ser concebidos para coletar, processar e disseminar (ao agente humano) informações relevantes sobre o ambiente ATFM, tais como demanda prevista, impactos projetados, variações na capacidade do espaço aéreo, etc. Em última instância, agentes inteligentes podem ser concebidos para gerar ações/políticas ATFM a partir da análise dos cenários aéreos, atuando diretamente no pro-

cesso decisório. Assim sendo, verifica-se que as pesquisas no referido domínio apresentam grandes desafios. Conforme discutido por Wolf (2007), quando considerados trabalhos empregando agentes, as questões apresentadas a seguir devem ser objeto de minuciosa análise.

- Modelagem de usuários e preferências: uma compreensão abrangente das entidades envolvidas é fundamental no processo de criação de agentes de apoio eficazes. A representação deve ser facilmente entendida e poderosa o suficiente para capturar os modelos de processo, as restrições ambientais e as intenções do usuário. Ou seja, os processos e os comportamentos humanos relevantes devem ser suficientemente compreendidos, de forma que possam ser imitados pelos agentes computacionais.
- Autonomia do agente computacional: qualquer tarefa, ao ser totalmente automatizada, pode implicar significativos ganhos para o usuário; contudo, devem ser tomados cuidados no processo de automatização das ações críticas; para tanto, um detalhado estudo deve definir quais tarefas podem ser executadas pelo agente de forma independente, sem comprometer a segurança, e como isso pode ser demonstrado.
- Atuação conjunta: existem tarefas que podem ser executadas somente pelo usuário, somente pelo agente ou em conjunto; embora a atuação conjunta seja, na maior parte das vezes, mais segura do que a atuação autônoma do agente computacional, podem surgir problemas de controle em ambientes onde as ações são compartilhadas; a cooperação é mais complexa do que uma atividade independente, na medida em que o agente computacional não deve contrariar as ações do agente humano.
- Aquisição de conhecimento: a implementação e/ou a adaptação de um sistema de agentes para a gestão do tráfego aéreo demanda conhecimento intensivo; alguns desses conhecimentos podem ser obtidos a partir de especialistas, mas nem sempre essa é uma possibilidade prática, ainda mais quando se verifica que determinado conhecimento não é dominado mesmo por especialistas.
- Aprendizagem a partir de dados históricos: quando disponíveis, os dados históricos podem servir como base para o desenvolvimento de modelos de comportamento humano, manualmente ou através de métodos de aprendizagem de máquina; as condições do ambiente/cenário e o *input* dos especialistas humanos devem formar o espaço de atributos, a partir do qual a decisão correta será aprendida.

- Adaptação às novas condições; assim como a aprendizagem pode ser usada para configurar o modelo inicial, ela também pode ser utilizada pelo agente para se adaptar/evoluir em ambientes de simulação; isto dá ao agente a capacidade de adaptar-se a situações diferentes daquelas encontradas nos dados históricos, bem como lhe permite corrigir erros no seu modelo.

A seguir, apresenta-se, em caráter introdutório, alguns trabalhos que ilustram algumas das linhas de pesquisa desenvolvidas no contexto ATFM.

3.1 Abordagem com Emprego de Programação Linear

O trabalho escolhido para referenciar esta linha de pesquisa foi o artigo denominado *A Stochastic Integer Program with Dual Network Structure and its Application to the Ground Holding Problem* (Ball et al, 2003). Neste trabalho, os autores propõem uma generalização do modelo clássico de fluxo em redes, mediante a substituição da demanda determinística por uma demanda estocástica. Na medida em que tal generalização destrói a estrutura original da rede, torna-se possível demonstrar que a matriz subjacente ao modelo estocástico constitui uma rede dual.

Desta forma, a programação inteira associada ao modelo estocástico pode ser eficientemente resolvida. Este modelo, tal como definido previamente, foi aplicado para solucionar o *Ground Holding Problem* (GHP), no domínio ATFM. Simplificadamente, pode-se dizer que os autores apresentaram uma função objetivo concebida para minimizar uma relação linear entre produção e demanda do fluxo sobre um modelo clássico de fluxo de redes, baseando-se em um conjunto de nós de origem e um conjunto de nós de destino (aeroportos). Este trabalho tem como foco, portanto, a proposição de um modelo para solucionar problemas de natureza demanda x capacidade em aeroportos.

A aplicação deste modelo no GHP, segundo os autores, agrega vantagens com relação a outras soluções propostas para tal problema, posto que possibilita uma fácil integração aos programas de gerenciamento de fluxo de tráfego, com espera no solo, baseados no paradigma do Processo Decisório Colaborativo (Collaborative Decision-Making paradigm).

3.2 Abordagem Empregando Programação Dinâmica e Teoria de Grafos

Pesquisadores ligados à Fundação Nacional de Ciência Natural da China propuseram um modelo baseado em grafos associado à Pesquisa Operacional (Ma *et al.*, 2004), onde a malha aérea foi representada por arcos e nós interligados, para cada arco associou-se um conjunto de atributos que restringiam o fluxo através dele e finalmente estabeleceu-se a função objetivo e as equações restritivas. Contudo, a computação da solução mostrou-se bastante demorada, tanto que Zhang, outro pesquisador chinês, propôs o melhoramento do trabalho mediante a utilização da técnica computacional conhecida como Programação Dinâmica, objetivando acelerar o processamento necessário para obter a solução.

Assim sendo, conforme pode ser verificado no artigo *Short-Term Flow Management Based on Dynamic Flow Programming Network* (Zhang *et al.*, 2005), o foco do modelo desenvolvido é a redução de congestionamentos nos setores de controle (e não nos aeroportos), mediante a utilização de esperas em solo e esperas em vôo. Segundo a concepção dos autores, a rede é compreendida como um grafo dirigido tri-dimensional $G(I, A)$, onde I é o conjunto de vértices e A é o conjunto de arcos dirigidos. Na rede, o aeroporto, os pontos de transferência e os auxílios à navegação são considerados vértices, que formam pares *source node-sink node*; os segmentos de aerovias e os corredores de aproximação são representados por arcos dirigidos.

O vôo flui através dos arcos, entre vértices, de forma ordenada e de acordo com as capacidades desses arcos. Assim sendo, a área onde se dará o controle de fluxo é representada por uma rede dinâmica tri-dimensional de fluxo. Neste contexto, o plano de vôo referente a uma aeronave corresponde a uma seqüência de arcos na rede do grafo. Uma aeronave executando uma espera em vôo (*airborn holding delay*) é considerada como um fluxo temporariamente escravizado a um nó correspondente a um ponto de transferência ou auxílio à navegação (*self-loop*). Uma aeronave executando uma espera em solo (*ground holding delay*) é considerada como um fluxo temporariamente escravizado a um nó correspondente a um aeroporto.

Tomando por base a concepção apresentada, o modelo para o gerenciamento do fluxo de tráfego aéreo pode ser considerado como um sistema aberto, que pode ser descrito como um problema de planejamento que leva em consideração os seguintes fatores: custos de combustível, custos da espera em solo e custos da espera em vôo. Para a implementação da solução via programação dinâmica,

são estabelecidas as restrições impostas pelas equações de conservação de fluxo e pelas capacidades dos arcos (segmentos de aerovias).

3.3 Abordagem Empregando Programação Inteira e Teoria de Grafos

A referência para a apresentação desta abordagem é o trabalho *The Air Traffic Flow Management Problem: An Integer Optimization Approach* (Bertsimas *et al*, 2008). Neste trabalho, os autores propõem a aplicação de programação inteira no problema de gerenciamento de fluxo de tráfego aéreo. O modelo proposto fornece uma representação completa de todas as fases de cada voo (decolagem, cruzeiro e pouso), e sugere ações para alcançar metas de segurança, eficiência e fluidez. A característica que diferencia o modelo é a capacidade de sugerir rotas alternativas.

O modelo matemático concebido é capaz de sugerir três tipos de ações, quais sejam: ajustar os horários de decolagem das aeronaves, ajustar a velocidade das aeronaves e modificar a rota das aeronaves. Este modelo foi baseado na proposta de Bertsimas e Pattersons (1998).

No modelo proposto, qualquer rota (origem-destino) é representada como uma seqüência de setores percorridos por uma aeronave. Nos modelos ATFM que não incluem a sugestão de rotas alternativas, a seqüência de setores a serem percorridos é pré-determinada. Para contemplar a mudança de rota no modelo matemático, o conjunto de setores que poderão ser percorridos deve ser aumentado.

Um elemento chave do modelo proposto é a definição de rotas. As rotas (origem-destino) podem ser representadas por grafos. O conjunto de nós do grafo (S_f) representa o conjunto de elementos do espaço aéreo dotados de capacidade, por exemplo, aeroportos e setores. O conjunto de arcos define as relações de seqüenciamento. Há um arco partindo de um nó i e chegando ao nó j se i e j são setores contíguos e o setor j pode ser voado logo após o setor i . Na Figura 3.3.1, três diferentes rotas entre o aeroporto de origem O_f e de destino D_f são descritas. No âmbito ATFM, pode-se supor, sem perda de generalidade, que o grafo de rotas $O_f - D_f$ é acíclico.

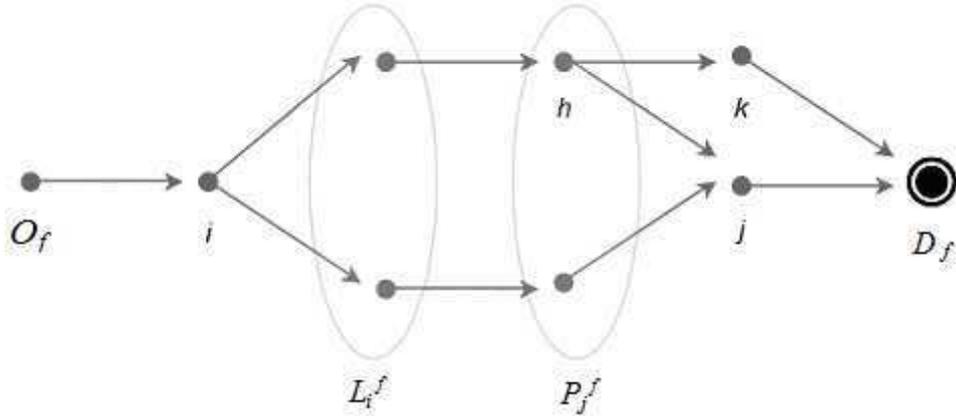


Figura 3.3.1: Grafo Origem - Destino

Tal condição permite dotar o conjunto de setores com uma relação binária e, portanto, imaginar o conjunto de rotas possíveis dentro de um conjunto parcialmente ordenado. O aeroporto de partida e o de chegada são os elementos mínimo e máximo (respectivamente) do conjunto parcialmente ordenado. O conjunto de rotas possíveis entre o par $O_f - D_f$ corresponde ao conjunto maximal de cadeias do conjunto parcialmente ordenado. Para impor que cada voo segua exatamente uma rota, estabelece-se uma condição local, qual seja: para voar em um setor, qualquer aeronave deve primeiro voar um dos setores anteriores por um período de tempo, no mínimo, equivalente ao tempo de voo nos referidos setores.

Com o intuito de formalizar esta condição para a definição de rotas, e conforme pode ser observado na figura 3.3.1, define-se a seguinte notação: para cada setor i ($\in S_f$), o subconjunto de setores subseqüentes a i é denotado por $L_i^f \subset S_f$. Analogamente, o subconjunto de setores que precede i é denotado por $P_i^f \subset S_f$.

Como na maioria dos modelos ATFM encontrados na literatura, o modelo proposto busca minimizar uma função de custo. Neste caso, a função de custo combina atrasos em voo (airborne-holding delay) e atrasos no solo (ground holding delay). Uma vez definida a função, são estabelecidas restrições para o cenário aéreo em função dos seguintes fatores: conjunto de aeroportos, conjunto de setores de controle, conjunto de setores que podem ser voados por determinado voo, conjunto de vôos, capacidade dos aeroportos, capacidade dos setores de controle, horários previstos de decolagem, horários estimados de chegada, origem dos vôos, destino dos vôos e tempo de percorrimento dos setores de controle.

De acordo com a configuração do cenário aéreo, definida a partir das restrições supracitadas, o modelo computa soluções que podem incluir esperas no solo, esperas em vôo e rotas alternativas.

3.4 Abordagem Empregando Técnicas Multi-agente

Por definição, o ATFM é uma atividade que tem a finalidade de assegurar um fluxo otimizado de tráfego aéreo para ou através de áreas onde demanda de tráfego temporariamente excede a capacidade disponível do sistema de controle de tráfego. Neste contexto, Dib (2005) apresentou um modelo com atuação na fase tática do ATFM, ou seja, a saída do sistema proposto consiste de ajustes nos horários de decolagens das aeronaves que estão com previsão de partir em até duas horas a contar do horário atual (do sistema).

Portanto, Dib (2005) apresenta um sistema para sincronização e gerenciamento de fluxo de tráfego aéreo, denominado de ATFM-MAS (*Air Traffic Flow Management Multi-agents System*), utilizando as tecnologias de Multi-Agentes e Grid Computacional. Desta forma, a partir da identificação da demanda esperada, resultante da análise das informações disponíveis dos horários de pousos e decolagens, havendo previsão de um congestionamento, os agentes entrarão em negociação, com o objetivo de ajustar o fluxo aéreo entre os aeroportos envolvidos.

A idéia é, baseando-se na possibilidade de se efetuar atrasos em determinadas decolagens, promover a concordância entre os aeroportos, de modo que seus agentes possam realizar a reprogramação dos horários, gerando soluções que reduzam o congestionamento, distribuindo o tráfego entre os aeroportos de forma mais uniforme.

Modelos da demanda esperada e dos recursos disponíveis são constantemente gerados, monitorados e atualizados pelo sistema. Para tanto, e considerando ainda a natureza distribuída da tarefa de gerenciamento do fluxo de tráfego aéreo, envolvendo múltiplos aeroportos, torna-se necessária a atuação de um conjunto de agentes, instalados nos diversos aeroportos. Os seguintes agentes foram propostos:

- Agentes ATC – tratam (em tempo real) as informações inerentes ao controle de tráfego aéreo (oriundas e destinadas aos controladores da TWR, APP e ACC), tais como o tempo de pouso das aeronaves.

- Agentes Pré-ATC – gerenciam o fluxo do aeroporto local (entrada/saída), com a função de manter uma escala de horários previstos para todos os vôos que saem e que chegam ao aeroporto, objetivando maximizar a utilização das pistas e possibilitar a identificação de congestionamentos.
- Agentes de Planejamento Tático – gerenciam o fluxo sobre todo o espaço aéreo, e monitoram os agentes Pré-ATC, as negociações entre os agentes, o fluxo de tráfego aéreo esperado para o aeroporto e do aeroporto; geram novas escalas de horários dos vôos, resultantes das negociações efetivadas, e distribuem tais escalas para os demais agentes envolvidos.

Os agentes negociam usando uma medida concreta, qual seja, o tempo de atraso por vôo, derivado do cálculo do Atraso Total de Peso Combinado (ATPC), definido como uma medida do fluxo e do congestionamento inerente à utilização dos recursos; o ATPC é uma função do tempo de atraso de cada vôo e do peso (fator de importância) a ele associado.

3.5 Abordagem Empregando Aprendizagem por Reforço

Nos últimos anos, a crescente demanda mundial para a utilização de aviões como meio de transporte tem aumentado consideravelmente a complexidade das atividades inerentes ao Gerenciamento do Tráfego Aéreo (ATM), entre as quais se inclui o Gerenciamento do Fluxo de Tráfego Aéreo (ATFM). Neste cenário, objetivando evitar ou reduzir os pontos de congestionamento do tráfego aéreo, são aplicadas medidas restritivas às aeronaves (no solo ou no ar). No entanto, tais medidas são concebidas e aplicadas por agentes humanos e os seus impactos, tanto operacionais (segurança) quanto financeiros, não são devidamente dimensionados, .

Para resolver este problema, Arruda (2009) propôs uma solução baseada em Aprendizagem por Reforço, desenvolvida segundo uma metodologia de HCI (*Human-Computer Interaction*). A solução foi modelada como um Sistema de Apoio à Decisão (SAD), através da construção do sub-módulo de modelagem e projeção de Impacto (MPI), parte do Módulo de Avaliação e Apoio à Decisão (MAAD), integrado ao Sistema de Apoio à Decisão Aplicado ao Gerenciamento Tático do Fluxo de Tráfego (SISCONFLUX).

A arquitetura de MAAD, segundo Arruda (2009), emprega agentes autônomos que, em situações de congestionamento dos setores de controle de tráfego aéreo, sugere medidas restritivas para ajustar o fluxo de tráfego (ground holding de-

lay). No trabalho proposto, o agente adquire o conhecimento do meio ambiente, através do algoritmo Q-learning, avaliando a situação do cenário aéreo que passa a existir após a aplicação da medida restritiva sugerida. A avaliação de cenário é feita através de uma função que computa os seguintes fatores: nível de congestionamento nos setores de controle, tempo de atraso imposto às aeronaves partindo de uma área específica, índice de impacto financeiro e fator de equidade na aplicação das medidas restritivas (atrasos nas decolagens).

Para tanto, houve a proposição de uma função de avaliação de cenário $G(z)$ composta por três termos, constando um indicador de congestionamento $C(z)$, um indicador de tempo de atraso nas terminais de controle $T(z)$ e um indicador de custo financeiro $I(z)$, com:

$$G(z) = \beta C(z) + \delta T(z) + \gamma I(z) \quad (3.5.1)$$

Onde os pesos β , δ e γ somados definem 100% do valor final de $G(z)$.

De forma mais específica, a proposta do trabalho em tela sugere a avaliação de impacto do congestionamento, em cada setor de controle, mediante a utilização da fórmula abaixo.

$$C_s(z) = \theta(Q(z) - c)\epsilon^{(Q(z)-c)} \quad (3.5.2)$$

Onde $C_s(z)$ se define a partir da quantidade de aeronaves $Q(z)$ que excede o limite de congestionamento c especificado para os setores de controle de tráfego aéreo; caso o resultado de $(Q(z) - c)$ seja negativo, a função de ajuste θ zera o valor retornado. O impacto total inerente ao congestionamento de todos os setores é definido por:

$$C(z) = \sum_{z=1}^n C_s(z) \quad (3.5.3)$$

No que tange ao dimensionamento do atraso (em solo e no ar) inerente a uma aeronave, o trabalho propõe a avaliação de impacto utilizando as expressões apresentadas a seguir.

$$d_{Ath}(a) = d_{Agh}(a) + d_{Aah}(a) \quad (3.5.4)$$

$$d_{Agh}(a) = \theta(t - (\alpha(t_{As}(a), t_{Aes}(a)))) \quad (3.5.5)$$

$$d_{Aah}(a) = \theta(t - (\alpha(t_{Ac}(a), t_{Aec}(a)))) \quad (3.5.6)$$

Onde:

- d_{Ath} representa o atraso total (em solo e no ar);
- d_{Agh} calcula o atraso no solo;
- d_{Aah} calcula o atraso em vôo;
- t é o horário atual;
- $t_{As}(a)$ é o horário programado para a saída da aeronave a ;
- $t_{Aes}(a)$ é o horário real de saída da aeronave a ;
- $t_{Ac}(a)$ é o horário programado para a chegada da aeronave a ;
- $t_{Aec}(a)$ é o horário real de chegada da aeronave a ;
- α - função que computa o tempo previsto ($t_{As}(a)$ ou $t_{Ac}(a)$), para os casos em que a aeronave a ainda não saiu do aeroporto ou ainda não chegou no aeroporto, ou seja, $t_{Aes}(a)$ e $t_{Aec}(a)$ indisponíveis;
- θ - função que retorna um resultado igual a 0, caso o valor de $Q(z) - c$ seja negativo.

Para definir o impacto imputado pelos atrasos em uma terminal de controle e no cenário como um todo, são empregadas as expressões apresentadas abaixo.

$$T_t(a) = \sum_{a=1}^n d_{Ath}(a) \quad (3.5.7)$$

$$T(z) = \sum_{z=1}^n T_t(z) \quad (3.5.8)$$

Onde:

- $T_t(a)$ - tempo de atraso das aeronaves por terminal de controle;
- $T(z)$ - indicador de atraso total no sistema.

Para avaliar o impacto em termos de equidade e custos financeiros, foi apresentada a expressão abaixo, composta por dois termos: indicador de equidade $J_{Ad}(a)$, e indicador de custo financeiro $F_{Ath}(z)$.

$$I(z) = J_{Ad}(z) + F_{Ath}(z) \quad (3.5.9)$$

O indicador utilizado para dimensionar o impacto referente à equidade se define por:

$$J_{Ad}(a) = \sum_{a=1}^n J(a) \quad (3.5.10)$$

Com:

$$J(a) = | 100 - (d_{Agh}(a).100/(d_{Td}/size(Term, t_{ini}, t_{fim}))) | \quad (3.5.11)$$

Onde:

- $d_{Agh}(a)$ - atraso no solo da aeronave a , em minutos;
- d_{Td} - atraso total inerente a uma terminal de controle, em minutos;
- $size$ - função que retorna a quantidade de aeronaves que sofreram atraso em um período de tempo;
- $Term$ - terminal onde o atraso ocorreu;
- t_{ini} - horário de início do cálculo de atraso;
- t_{fim} - horário final do cálculo de atraso.

O indicador utilizado para dimensionar o impacto referente ao custo financeiro se define por:

$$F_{Ath}(a) = \theta(d_{Ath}(a) - D_a)^{1+\theta(d_{Ath}(a)-C_a)} \quad (3.5.12)$$

Onde:

- C_a - variável definida em função do tempo máximo que um vôo pode atrasar, antes de ser cancelado;
- D_a - variável definida em função do tempo máximo que um vôo pode atrasar, sem prejudicar outros vôos;
- $d_{Ath}(a)$ - atraso total da aeronave a , em minutos;
- θ - função que retorna um resultado igual a 0, caso o valor seja negativo

Na proposta em tela, as equações apresentadas compuseram a função de avaliação utilizada na estrutura de recompensa do Algoritmo Q-learning, utilizado no desenvolvimento do protótipo que foi testado. A aplicação do modelo proposto tornou possível a geração de políticas de controle de fluxo de tráfego baseadas nos critérios supracitados.

A proposta contemplou a realização de estudo caso incluindo o teste de protótipos configurados de duas formas distintas: protótipo utilizado os critérios

de congestionamento, atraso e equidade/financeiro; protótipo utilizando apenas os critérios de congestionamento e atraso. Os resultados obtidos indicaram que a utilização do critério equidade/financeiro implica em montantes de atrasos maiores nas terminais de controle, decorrendo importante dilema entre custo financeiro e fluidez/segurança na atividade ATFM.

3.6 Estado da Arte

No domínio ATFM, os trabalhos mais recentes levantados combinam formalismos de Aprendizagem por Reforço e Técnicas Multi-Agentes. No trabalho de Agogino e Tumer (2009) intitulado *Learning Indirect Actions in Complex Domains: Actions Sugestions for Air Traffic Control*, técnicas multi-agentes associadas a algoritmos de aprendizagem por reforço são utilizados para agregar a experiência do especialista (agente humano) a processos ATFM automatizados.

Conforme levantado por Agogino e Tumer no trabalho supracitado, os domínios relacionados ao controle de tráfego aéreo são demasiadamente complexos para serem completamente automatizados e, de fato, o ATFM tático tem impacto direto sobre as atividades ATC. A Figura 3.6.1 apresenta o fluxo de informações/ações considerando um sistema onde o agente atua de forma totalmente automatizada.

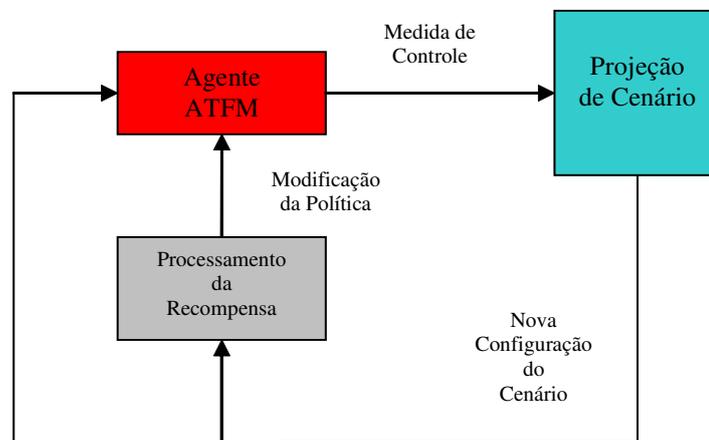


Figura 3.6.1: Sistema Totalmente Automatizado (Agogino e Tumer, 2009)

No processo totalmente automatizado (Figura 3.6.1), a medida de controle sugerida pelo Agente ATFM é aplicada diretamente, modificando o fluxo de tráfego. Então, o cenário modificado é avaliado segundo os critérios pré-estabelecidos e o sistema computa uma recompensa para o referido agente. Esta recompensa, então, é utilizada para modificar a política de controle adotada pelo mesmo, o que implicará modificação na próxima medida sugerida.

No domínio ATFM, contudo, a configuração mais adequada sugere um processo decisório onde um agente computacional oferece sugestões a um agente humano, que atua como um filtro ao acatar ou rejeitar a sugestão do agente computacional. A figura 3.6.2 ilustra tal concepção.

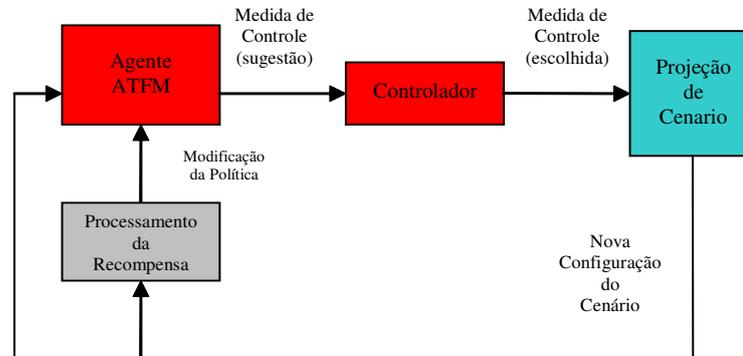


Figura 3.6.2: Sistema Semi-automatizado

No estudo em tela, os autores desenvolveram métricas para avaliar a eficiência de medidas de controle de fluxo de tráfego geradas por agentes baseados em algoritmos de aprendizagem por reforço (Agente ATFM). Como resultado, foi possível aprimorar as ações de um agente concebido para simular o comportamento dos controladores de tráfego aéreo (Agente Controlador), em função das sugestões oferecidas pelo Agente ATFM, e na medida em que tais sugestões eram acatadas pelo Agente Controlador. A Figura 3.6.3 ilustra o modelo proposto por Agogino e Tumer.

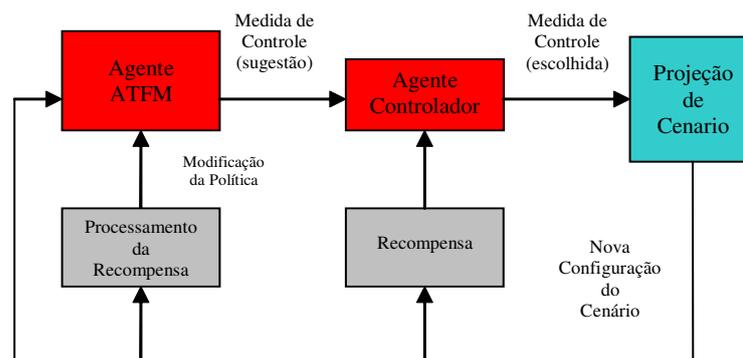


Figura 3.6.3: Sistema Semi-automatizado (Agogino e Tumer, 2009)

Neste ponto, cabe ressaltar que no sistema proposto por Agogino e Tumer, o Agente ATFM é um algoritmo de aprendizagem por reforço buscando maximizar a recompensa derivada dos cenários gerados, que serão avaliados em função do montante de atraso produzido e da quantidade de aeronaves nos setores de controle, ambos decorrentes da política de controle de fluxo adotada.

Por sua vez, o Agente Controlador é caracterizado como um agente (também baseado em um algoritmo de aprendizagem por reforço) maximizador de recompensa que busca apenas a redução da quantidade de aeronaves nos setores de controle. Neste ponto, cabe ressaltar que esta concepção do comportamento do controlador de voo constitui uma premissa de validade restrita, na medida em que a redução de aeronaves nos setores, apesar de implicar a diminuição da carga de trabalho dos controladores e tornar a operação ATC mais segura, não constitui o único objetivo da atividade ATC, a qual também busca a máxima fluidez de tráfego possível. Tal condição portanto constitui um importante limitador das simulações apresentadas, posto que, na prática, o controlador humano é um agente que busca a otimização do fluxo e a segurança nas operações, ressaltando que esta última prepondera de forma indiscutível.

No processo representado pela Figura 3.6.3, o Agente ATFM sugere uma medida de controle e o Agente Controlador tem a opção de aceitá-la ou rejeitá-la. Na proposta dos autores, o Agente Controlador possui uma estrutura de recompensa concebida para incentivar a aceitação das sugestões. Ao efetivar a sugestão, o novo cenário, então, é projetado e os valores referentes ao montante de atraso e a quantidade de aeronaves nos setores de controle são computados. Com base nesses valores, o sistema computa o valor da recompensa e, como consequência, as políticas (ações) futuras serão impactadas.

Como resultado das simulações efetuadas, mediante a utilização de dados reais, os autores verificaram que a atuação conjunta do Agente ATFM com o Agente Controlador (sugestões acatadas) apresentou uma redução de congestionamentos da ordem de 20% quando comparadas a atuação isolada do Agente Controlador (sugestões rejeitadas).

Capítulo 4

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo são discutidos os fundamentos teóricos que embasam a aplicação dos formalismos da Inteligência Artificial empregados na proposta apresentada neste trabalho.

4.1 Processo Decisório

Conforme colocado por Wickens e Hollands (1999), diversos acidentes graves, envolvendo erros humanos, foram atribuídos à tomada de decisão dos operadores. A complexidade de uma escolha, no contexto de um processo decisório, pode variar muito dependendo de fatores como:

- a) Incerteza: o grau de incerteza das conseqüências de uma decisão e seus riscos, consideradas as situações onde não existe um meio sistematizado de projetar e avaliar os impactos da mesma.
- b) Experiência e Habilidade: as decisões, tomadas por pessoas experientes e com habilidade certificada, podem surgir mais rapidamente e com menos esforço. Contudo, essa condição não garante que tais decisões sejam as mais adequadas, mormente quando o domínio trabalhado tem natureza complexa, o que se caracteriza, via de regra, pela necessidade de se analisar cenários definidos a partir de diversos fatores e variáveis.
- c) Tempo: a premência requerida na adoção de determinadas medidas tem implicação direta na qualidade de decisão tomada. Em ambientes e cenários analisados em tempo real, ou próximo de tal condição, o tempo deve ser considerado um fator crítico.

4.1.1 Modelo de Processamento de Informação para Tomada de Decisão

Segundo Wickens e Hollands (1999), o processo inerente a tomada de qualquer decisão pode ser modelado segundo passos bem definidos, a saber.

- **1º Passo: Detecção da Situação**

O agente tomador de decisão precisa buscar sugestões ou informações sensoriais do ambiente. Estas sugestões são sempre confusas, incertas, ambíguas e podem ser interpretadas incorretamente. A atenção seletiva possui uma função muito importante na tomada de decisão e na seleção de sugestões através de um filtro. Esta seleção é baseada em experiências passadas (memória de longo prazo) e requer mais esforço. As sugestões são então selecionadas e percebidas através do entendimento e verificação da situação levando ao diagnóstico. A combinação das operações de percepção, memória ativa e cognição que faz com que o tomador de decisão possa criar hipóteses sobre o estado atual e futuro do mundo, este processo chama-se verificação / conhecimento da situação.

- **2º Passo: Diagnóstico**

O diagnóstico é baseado em duas fontes de informação: as sugestões externas filtradas pela atenção seletiva e a memória de longo prazo. A memória de longo prazo pode trazer ao tomador de decisões várias hipóteses do estado atual do mundo. Muitos diagnósticos são iterativos, isto é, a hipótese inicial pode disparar outras informações que podem confirmar ou rejeitar a hipótese inicial.

- **3º Passo: Escolha da Ação**

A partir da memória de longo prazo, o agente tomador de decisão pode gerar um conjunto de ações ou opções de decisões; mas se o diagnóstico é incerto, as consequências das diferentes escolhas podem definir os riscos. A consideração a respeito dos riscos requer uma estimativa de valores. A decisão é verificada para fins de confirmação / rejeição, aprendizado, melhoria de decisões futuras. Este retorno é armazenado na memória de longo prazo a fim de que o tomador de decisão possa revisar as suas regras e estimar melhor os riscos. A meta-cognição é o conhecimento sobre as limitações do tomador de decisão a respeito de suas próprias decisões.

4.1.2 Qualificação da Decisão

O objetivo do agente tomador de decisão é, via de regra, proporcionar algum tipo de benefício para um grupo, comunidade ou sistema. Assim sendo, a qualificação de uma decisão se fundamenta em algumas premissas, tal como especificado a seguir.

- Valor esperado: uma boa decisão deve produzir o máximo valor de retorno, o que se materializa em termos de benefícios para um sistema (por exemplo). Entretanto, este valor depende de critérios e parâmetros que devem ser previamente estabelecidos e, como consequência, deve ser implementada uma metodologia que permita avaliar determinado sistema segundo tais critérios e parâmetros.
- Retorno: este fator constitui uma especificidade do Valor Esperado, e deve assegurar que boas decisões produzam bons resultados e decisões ruins produzam resultados ruins.
- Habilidade: este fator representa a capacidade do agente tomar boas decisões em função de determinadas condições, tais como conhecimento do ambiente, experiência prévia, capacidade de aprendizagem, etc. A preparação (ou concepção) de um agente tomador de decisão deve levar em consideração tais condições.

4.1.3 Diagnóstico e Detecção da Situação

O entendimento da situação é um dos componentes mais efetivos na tomada de decisão. Existem componentes do processamento de informação que podem influenciar na qualidade do diagnóstico:

- A Percepção das Sugestões: as sugestões são probabilísticas e requerem alguma inferência para gerar uma situação real. Os humanos são relativamente eficientes para estimar valores médios, proporção, entretanto, com proporções extremas, tendem a ser conservadores. Para seres humanos, a variabilidade é melhor percebida quando se compara valores menores. Com relação às tendências, os seres humanos extrapolam tendências não lineares. Como consequência, percebe-se que o tomador de decisão humano enfrentará progressiva dificuldade em lidar com cenários que agreguem um número crescente de variáveis.
- A Atenção para Seleção e Integração da Informação: este processo envolve múltiplas sugestões e integração de diversas informações. Considerando-se o agente humano, uma maior quantidade de informação não leva necessariamente a decisões mais precisas, na medida em que as pessoas tendem a ignorar a maior parte das informações mediante a aplicação de filtros, o que dificulta a adequada integração das referidas informações e, conseqüentemente, verifica-se uma degradação no processo de compilação da situação (cenário). Em adição, informações que requerem cálculos matemáticos são difíceis de interpretar, integrar e em alguns casos são

ignorados. Reforça-se, portanto, a necessidade de agregar, ao processo decisório, agentes computacionais que apoiem o agente humano.

- A Memória de Longo Prazo (experiência): o agente humano diagnostica uma situação, escolhe uma hipótese, percebe uma evidência com base na experiência, isto é, na memória de longo prazo. Caso a evidência de uma determinada situação seja ambígua, o diagnóstico será feito com base na experiência. Quanto maior a disponibilidade e mais recente for a informação na memória de longo prazo, maior a probabilidade de a informação ser considerada na tomada de decisão. O excesso de confiança, proveniente da memória de longo prazo, pode fazer com que a busca da informação correta não seja feita. Verifica-se que a mente humana tende a atribuir maior valor à primeira impressão, isto é, ao primeiro registro de memória encontrado (correlacionado à decisão a ser tomada ou ao fator a ser analisado), o que pode levar a um erro na decisão pela busca de confirmações da informação, mesmo que esta esteja errada. Neste ponto, ressalta-se os ganhos que podem ser obtidos através da incorporação da experiência do *expert* humano por agentes computacionais engajados no suporte ao processo decisório.

4.1.4 Escolha da Ação

Todo o processamento de informações é necessário para sustentar a escolha da ação. A satisfação na escolha da ação está associada à melhor escolha com base em um conjunto de atributos, definidos por ordem de importância. A incerteza na escolha da ação pode ser resultado da falta de conhecimento e/ou compreensão da situação atual (cenário). Outro fator importante na tomada de decisão é o gasto ou economia de esforço para executar uma tarefa, isto é, a decisão está associada à antecipação de esforço a ser despendido.

A escolha por situações arriscadas (exemplo: não conformidade com processos de segurança e qualidade), está associada à certeza do "custo da conformidade"; então, o agente pode preferir assumir o risco adotando uma ação que demande menor esforço. Neste caso, o que se interpreta é que o custo e esforço a ser empregado para obter a conformidade é menor que o risco da não-conformidade. O conhecimento e entendimento sobre o problema, o esforço da ação e o resultado esperado antecipam a tomada de decisão e a implementação da ação. Outro aspecto importante é o monitoramento constante da decisão e escolha da ação (constante dimensionamento dos possíveis impactos da ação). Neste contexto, pode-se destacar, de forma resumida, alguns pontos que podem melhorar a capacidade de tomar decisões:

- Habilidade e experiência;
- *Feedback* sobre decisões e ações;
- Treinamento em processamento de múltiplas informações, fontes e pressão de tempo;
- Elaboração de procedimentos que devem ser seguidos para melhorar a qualidade da tomada de decisão (*checklist*); e
- Automação.

A partir das considerações apresentadas acima, e considerando a proposta do presente trabalho, torna-se clara a necessidade de se utilizar formalismos que tornem possível incorporar, a um sistema computacional, a capacidade de atuar em um processo decisório específico, dadas as características e especificações inerentes ao mesmo. Assim sendo, apresenta-se a seguir formalismos estudados pela Inteligência Artificial com possibilidade de implementar tal capacidade.

4.2 Teoria da Decisão

4.2.1 Teoria da Utilidade e Tomada de Decisão Simples

Uma determinada ação pode ser adotada ou rejeitada na medida em que atinja (ou não) metas estabelecidas, independentemente do fato de que outras ações poderiam levar ao cumprimento da referida meta. Contudo, quando o cenário de decisão agrega algum grau de incerteza no que tange ao resultado de tal ação, o agente necessita de suportes adicionais para decidir entre ações possíveis. Assim sendo, para realizar a escolha adequada, o agente deve ter preferências entre os diferentes resultados possíveis, definidos como estados (cenários) completamente caracterizados (Russel, 2003).

A Teoria da Utilidade permite trabalhar com preferências, ao estabelecer que todo estado tem um determinado grau de utilidade para um agente específico, e tal agente preferirá estados com utilidade mais alta. Portanto, conforme Russel e Norvig (2003), a utilidade de um estado é específica para o agente cujas preferências a função utilidade foi concebida para representar; considerando ainda que tal função gera um único número para expressar a desejabilidade do referido estado.

Partindo do caso básico mais geral, onde os estados são observados como instantâneos completos do mundo, e onde uma ação não determinística A terá

estados resultantes possíveis $resultado_i(A)$, onde o índice i varia sobre diferentes resultados, tem-se que, antes da execução de A , o agente atribui a probabilidade $P(resultado_i(A) | Fazer(A), E)$ a cada resultado, dado que E resume a evidência disponível ao agente sobre o mundo e $Fazer(A)$ é a proposição de que a ação A é executada no estado atual. Calcula-se, então, a utilidade esperada da ação dada a evidência, $UE(A | E)$.

$$UE(A | E) = \sum_i P(resultado_i(A) | Fazer(A), E) \cup (resultado_i(A)) \quad (4.2.1)$$

No âmbito da Teoria da Utilidade, surge o conceito da Utilidade Máxima Esperada (UME), segundo a qual um agente é racional se e somente se ele escolhe a ação que resulta na mais alta utilidade esperada, calculada como a média sobre todos os resultados possíveis da ação. Este princípio define a ação certa a ser adotada em qualquer problema de decisão. Contudo, para escolher a melhor seqüência de ações usando a equação acima, seria necessário enumerar todas as seqüências de ações, o que se torna inviável para seqüências longas. Tais seqüências serão tratadas mais adiante. Neste ponto, o foco recairá sobre ações únicas (decisões simples). O arcabouço epistemológico que embasa a Teoria da Utilidade abarca seis restrições sobre preferências racionais conhecidas como axiomas da Teoria da Utilidade, os quais especificam as restrições semânticas mais óbvias sobre preferências e loterias (Russel e Norvig, 2003). Isto posto, e considerando que a preferência é uma propriedade básica de agentes racionais, a partir dos axiomas de utilidade torna-se possível definir uma função utilidade, estabelecida segundo o Princípio da Utilidade.

O enunciado do Princípio da Utilidade estabelece que, se as preferências de um agente obedecem aos axiomas de utilidade, então existe uma função de valores reais U que opera sobre estados (A e B e por exemplo) tais que $U(A) > U(B)$ se e somente se A é preferível em relação a B , e $U(A) = U(B)$ se e somente se o agente está indiferente entre A e B . Como conseqüência, a fórmula da utilidade esperada pode ser escrita da seguinte forma:

$$U([p_1, S_1; \dots; p_n, S_n]) = \sum_i p_i U(S_i) \quad (4.2.2)$$

Onde $U([p_i, S_i; \dots; p_n, S_n])$ é a utilidade da loteria $[p_i, S_i; \dots; p_n, S_n]$ (probabilidade p_i de que ocorra como resultado o estado S_i). Considerando aplicações onde o agente atua em ambientes determinísticos, utiliza-se uma particularização da Função Utilidade denominada Função de Valor ou Função de Utilidade ordinal, conforme definido abaixo.

$$U(S) = k \quad (4.2.3)$$

Contudo, para as aplicações propostas neste trabalho, emerge uma questão fundamental, qual seja, a necessidade de se avaliar diversos atributos em um determinado estado. De maneira mais específica, o desenvolvimento do MAAD requer a construção de uma função utilidade que abarque todos os critérios de avaliação de cenários definidos pelo usuário do sistema.

A Teoria de Utilidade Multi-atributo (Russel, 2003) apresenta a solução para essa questão. Considerando os atributos de $X = X_1, \dots, X_n$ e o vetor completo de atribuições $x = \langle x_1, \dots, x_n \rangle$, e supondo que cada um dos atributos admita valores escalares (discretos ou contínuos) distintos possíveis, no pior caso, para especificar a função utilidade completa seriam necessários d^n valores. E o pior caso decorre de uma condição segundo a qual as preferências do agente não apresentam nenhuma regularidade.

A Teoria de Utilidade Multi-atributo, contudo, está fundamentada na suposição de que as preferências de agentes típicos são de alguma forma estruturadas. Uma vez identificadas regularidades no comportamento das preferências e aplicando-se teoremas de representação (Evans e Garipey, 1992), pode-se definir uma estrutura de preferências e utilidade multi-atributo, conforme especificado a seguir.

$$U(x_1, \dots, x_n) = f[f_1(x_1), \dots, f_n(x_n)] \quad (4.2.4)$$

Em complemento, segundo o princípio da Independência Preferencial Mútua (Teorema de Debreu), pares de atributos são preferencialmente independentes em relação aos demais. Assim sendo, o comportamento preferencial do agente pode ser descrito como uma maximização da função de valor multi-atributo definida por:

$$V(x_1, \dots, x_n) = \sum_i V_i(x_i) \quad (4.2.5)$$

A Teoria da Decisão, por sua vez, emerge quando as preferências (utilidades) são combinadas com probabilidades. Ou seja, a Teoria da Decisão resulta de uma combinação da Teoria da Utilidade com a Teoria da Probabilidade. Em suma, segundo a Teoria da Decisão, um agente é racional se e somente se ele escolhe a ação que resulta na mais alta utilidade esperada, definida a partir da média entre todos os resultados possíveis da ação.

4.2.2 Tomada de Decisão Complexa (Decisões Seqüenciais)

Uma vez analisado o processo de tomada de decisões instantâneas (ou episódicas), passa-se à análise do problema de decisão seqüencial, onde a utilidade do agente depende de uma seqüência de decisões. O modelamento deste problema, a partir de um estado inicial S_0 , requer a especificação de uma função (ambiente determinístico) ou modelo (ambiente estocástico) de transição de estados e uma função de recompensa. Neste contexto, a especificação da função utilidade dependerá de uma seqüência de estados (histórico de ambiente).

Considerando um ambiente completamente observável, onde as transições entre estados observam as propriedades de Markov e utiliza-se recompensas aditivas, o problema de decisão seqüencial se caracteriza como um Processo de Decisão de Markov (MDP). O MDP será melhor explicado posteriormente, mas, considerando o problema de Decisões Seqüenciais, um modelo de transição de estado $T(s, a, s')$ probabilisticamente especifica o próximo estado do ambiente (s') em função do seu estado atual (s) e da ação do agente (a). Uma função de recompensa $R(s)$ especifica a recompensa instantânea esperada em função do estado atual e da ação. Em suma, o problema de decisão seqüencial constitui um modelo de Markov se as transições de estado são independentes de quaisquer estados anteriores do ambiente ou ações anteriores do agente.

Conforme especificado acima, o modelo de transição $T(s, a, s')$ realiza o mapeamento dos efeitos da ação a sobre o estado (cenário) s , resultando o novo estado s' . A função de recompensa $R(s)$, por sua vez, é constituída por uma função utilidade aditiva, melhor explicada mais adiante, com valor definido em função das características do estado s .

Quando se trata de decisões seqüenciais, é fundamentalmente importante definir se existe um horizonte finito ou um horizonte infinito para a tomada de decisão. A constatação de que o horizonte é finito implica a existência de um tempo fixo N , após o qual não há mais sentido tomar decisões, pois o processo decisório foi encerrado. Tem-se, portanto que:

$$U_h([s_0, s_1, \dots, s_{N+k}]) = U_h([s_0, s_1, \dots, s_N]), \text{ para todo } k > 0 \quad (4.2.6)$$

Como consequência, pode-se inferir que, com um horizonte finito, a ação ótima (comportamento do agente) para um determinado estado pode se alterar em função do tempo, decorrendo que a política ótima, no contexto de um

processo decisório com horizonte finito, é não-estacionária. Nos processos decisórios com horizonte infinito, o agente deve se comportar da mesma maneira ao defrontar estados iguais, mesmo que tal condição ocorra em tempos diferentes. Conseqüentemente, a ação ótima depende apenas do estado atual, e a política ótima é definida como estacionária.

No processo decisório envolvendo decisões complexas, o cálculo da utilidade de seqüências de estados pode ser implementado com o auxílio da teoria da utilidade multi-atributo, onde cada estado s_i é visualizado como um atributo da seqüência de estados $[s_0, s_1, s_2, \dots, s_{i-1}]$. Contudo, a definição de uma expressão para avaliação da utilidade de seqüência de estados requer algum tipo de suposição em termos de independência das preferências. E a suposição mais natural é que as preferências do agente, entre seqüências de estados, são estacionárias. E o caráter estacionário para preferências implica que, se duas seqüências de estados $[s_0, s_1, s_2, \dots]$ e $[s'_0, s'_1, s'_2, \dots]$ começam com o mesmo estado ($s_0 = s'_0$), então as duas seqüências devem ser ordenadas por preferência do mesmo modo que as seqüências $[s_1, s_2, \dots]$ e $[s'_1, s'_2, \dots]$. Dado o caráter estacionário, decorrem apenas duas maneiras de atribuir utilidades a seqüências:

Recompensas Aditivas

$$U_h([s_0, s_1, s_2, \dots]) = R(s_0) + R(s_1) + R(s_2) + \dots \quad (4.2.7)$$

Recompensas Descontadas

$$U_h([s_0, s_1, s_2, \dots]) = R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \dots, 0 < \gamma \leq 1 \quad (4.2.8)$$

O fator de desconto γ define a preferência do agente entre recompensas atuais e futuras. Quando γ é próximo de 0, as recompensas no futuro distante tornam-se insignificantes. Ao passo que quando γ é 1, as recompensas descontadas tornam-se aditivas. A adoção de recompensas descontadas permite a atribuição de uma utilidade finita a uma seqüência infinita, mediante a utilização da fórmula para a soma de uma série geométrica infinita, sempre que as recompensas são limitadas por R_{max} e $\gamma < 1$:

$$U_h([s_0, s_1, s_2, \dots]) = \sum_{t=0}^{\infty} \gamma^t R(s_t) \leq \sum_{t=0}^{\infty} \gamma^t R_{max} = R_{max}/(1 - \gamma) \quad (4.2.9)$$

Para o caso específico do MAAD, o ambiente comporta estados terminais, que se caracterizam pela condição em que o cenário atende às especificações

(parâmetros e critérios) estabelecidas pelo usuário, não sendo requeridas ações adicionais. Ou seja, mediante a adoção de uma política adequada, o agente assegura que estados terminais são alcançados, o que permite a utilização de recompensas aditivas ou descontadas.

4.3 Processo Decisório de Markov

O arcabouço matemático que embasa a AR é conhecido como Processo Decisório de Markov (MDP), que é caracterizado a partir da Propriedade de Markov, segundo a qual estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido. De forma geral, um Processo de Markov $\{X_n\}$ é um processo estocástico que, considerado o valor X_n , os valores X_p (para $p > n$) não são influenciados pelos valores de X_m , para $m < n$, com $m, n, p \in \mathbb{N}$. Formalmente, portanto, apresenta-se o enunciado da Propriedade de Markov, dado um conjunto de estados S :

$$Pr \{X_{n+1} = j / X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n\} = Pr \{X_{n+1} = j / X_n = i_n\} \quad (4.3.1)$$

para todo $n \in \mathbb{N}$ e para todos os estados $i_0, \dots, i_{n-1}, i_n, j \in S$.

Considerando a satisfação da Propriedade de Markov, o Processo Decisório de Markov se define a partir de: a) um conjunto de estados possíveis (S); b) um conjunto de ações possíveis (A); c) uma função de recompensa ($R(s)$ ou $R(s, a)$); e d) uma descrição do modo como as ações impactam cada estado (T) (Suton e Barto, 1998). A partir das condições estabelecidas acima, então, parte-se para a busca de um conjunto de ações (política de ações π) que maximize as recompensas obtidas pelo agente, fazendo com o que o mesmo atinja o(s) objetivo(s) para o qual foi concebido. Formalmente, um PDM padrão é definido como se segue:

- Conjunto de estados S com um estado inicial s_0 ;
- Conjunto de ações A ;
- Função de recompensa $R(s)$, definida por $R : S \times A \rightarrow \mathfrak{R}$
- Modelo de transição: $T(s, a, s')$, definido por $T : S \times A \rightarrow \Pi(S)$, onde cada elemento do conjunto $\Pi(S)$ representa uma distribuição de probabilidade sobre o conjunto S .

Desta feita, em um ambiente onde o resultado no momento $t + 1$, para uma ação praticada em t , depende de todo o histórico de ações até o momento atual

(a dinâmica do ambiente é definida pela especificação completa da distribuição de probabilidades), conforme apresentado abaixo, não se caracteriza um PDM.

$$Pr \{s_{t+1} = s', r_{t+1} = r \mid s_t, a_t, r_t, s_{t-1}, a_{t-1}, r_{t-1}, \dots, s_0, a_0, r\} \quad (4.3.2)$$

onde Pr é a probabilidade do estado s_{t+1} ser o estado s' , considerada uma função que depende de todos os estados, ações e recompensas passadas.

Por outro lado, se a resposta do ambiente em $t + 1$ depende apenas do estado (vigorando) e da ação (praticada) em t (a dinâmica do ambiente satisfaz a Propriedade de Markov), caracteriza-se um PDM, onde a probabilidade da transição para o estado s' , a partir da execução da ação a , expressa-se por:

$$P_{s,s'}^a = Pr \{s_{t+1} = s' \mid s_t = s, a_t = a\} \quad (4.3.3)$$

4.4 Aprendizagem por Reforço

A Aprendizagem por Reforço (AR) é uma vertente metodológica da Inteligência Artificial que permite agregar a um agente (sistema, módulo, etc.) a capacidade de perceber determinado ambiente (cenário, realidade, etc.), de atuar nesse ambiente ou apresentar um comportamento específico em função desse ambiente e, em função das conseqüências da sua atuação, receber uma recompensa ou uma punição em termos numéricos. Mediante tentativa e erro, o agente aprende a agir de forma a maximizar as recompensas (Kaelbling *et al*, 1996). Ou seja, o cerne da Aprendizagem por Reforço consiste em se definir uma política de ações que maximize o total de recompensas recebidas pelo agente.

No modelo de Aprendizagem por Reforço padrão, um agente está ligado ao seu meio ambiente (cenário) através da percepção e da ação, conforme representado na Figura 4.4.1. Em cada passo da interação, o agente recebe como entrada (i) uma indicação do estado atual (s) do ambiente; o agente escolhe então uma ação (a) para gerar como saída. A ação muda o estado do ambiente, e o valor desta transição de estado é comunicado ao agente através de um sinal de reforço escalar (r). O comportamento do agente (B) deve escolher ações que tendem a aumentar, a longo prazo, a soma dos valores do sinal de reforço. Ele pode aprender a fazer isso, ao longo do tempo, por tentativa e erro, guiado por uma grande variedade de algoritmos. Formalmente, o modelo consiste de:

- um conjunto discreto de estados S ;
- um conjunto discreto de ações A que podem ser executadas pelo agente;

e

- um conjunto de sinais de reforço.

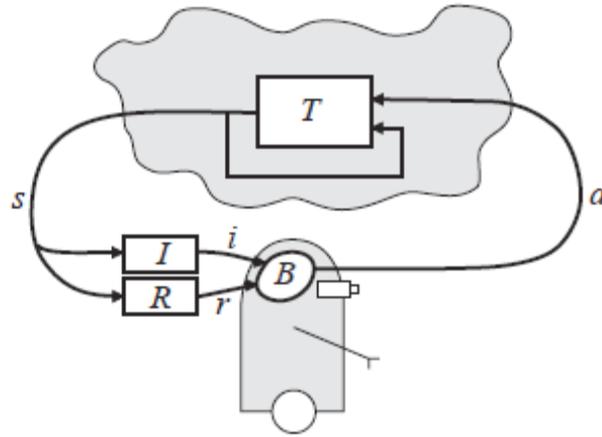


Figura 4.4.1: Modelo Padrão de Aprendizagem por Reforço (Kaelbling *et al.*,1996)

A Aprendizagem por Reforço difere, em diversos pontos, das formas de aprendizagem supervisionada mais amplamente estudadas. A diferença mais importante é que não existem informações prévias inerentes a pares de *input / output*. De fato, o que ocorre é que o agente, após escolher sua ação, recebe a recompensa imediata e informações sobre o estado subsequente; contudo, não há informação sobre qual ação teria sido melhor para seus interesses de longo prazo. Desta feita, é necessário que o agente, para agir de forma otimizada, adquira experiência sobre estados possíveis, ações, transições e recompensas. Outra diferença relevante recai sobre a importância (para a AR) do desempenho *on line* do sistema, na medida em que o processo de avaliação de estados e o processo de aprendizagem são frequentemente concomitantes (Kaelbling *et al.*,1996).

Por estado, entende-se um conjunto de características indicando como está o ambiente (ou cenário), o que se configura a partir das percepções do agente e da forma como o universo de atuação (do agente) foi modelado. As informações sobre o estado devem permitir ao agente definir ações a serem executadas; ou seja, a representação desse estado deve ser suficiente para que o referido agente tome suas decisões, satisfazendo a propriedade de Markov, de modo que a decisão sobre a ação a ser tomada não dependa da seqüência de estados anteriores.

A função de recompensa gera o *feedback* do agente em função do comportamento (ação) do agente, onde $R(s, a)$ representa o retorno recebido após a execução de uma ação a ($a \in A$) a partir de um estado s ($s \in S$). A função

recompensa é definida por $r : (S \times A) \rightarrow \mathbb{R}$. Já a descrição de como as ações impactam cada estado se define a partir de uma função (ambientes determinísticos) ou de um modelo (ambientes não determinísticos) de transição de estados. A função $f : (S \times A) \rightarrow S$ (ou modelo $\delta : (S \times A) \rightarrow S$) indica qual é o estado em vigor considerando que ao estado s foi aplicada a ação a . O modelo de transição de estados $\delta(s, a, s')$ indica a probabilidade de vigorar um estado s' , dado que a um estado s foi aplicada uma ação a .

A política de ações para um conjunto de estados $\pi(s)$ se define por $\pi : S \rightarrow A$, ou seja, representa a função que modela o comportamento do agente, através do mapeamento estado-ação. A política $\pi(s)$ pode ser vista como um conjunto de regras do tipo $s_n \rightarrow a_m$. Tomando por base o que foi apresentado em 4.2, verifica-se que o agente pode avaliar se determinado estado é bom ou ruim (utilidade). Tal capacidade decorre do estabelecimento de uma Função Valor de Estados definida a partir das recompensas e da política de ações. Esta função valor representa a recompensa inerente a um estado s acrescida das recompensas futuras derivadas da adoção da política π , e se define por $V^\pi(s) : S \rightarrow \mathbb{R}$. Tem-se que:

$$V^\pi(s_0) = r_0 + r_1 + r_2 + \dots \quad (4.4.1)$$

Tal como visto anteriormente, o caso em que as ações do agente ocorrem de forma indefinida (tempo infinito), a função valor tende a valores infinitos. desta feita, visando garantir convergência e diferenciar recompensas distantes (temporalmente) do estado atual, aplica-se um fator de desconto $0 \leq \gamma \leq 1$, decorrendo:

$$V^\pi(S_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \dots \quad (4.4.2)$$

$$V^\pi(S_t) = r_t + \gamma V^\pi(S') \quad (4.4.3)$$

Onde:

$$r_t = r(S_t, \pi(S_t)) \quad (4.4.4)$$

$$S' = \delta(S_t, \pi(S_t)) \quad (4.4.5)$$

De forma análoga, pode-se definir uma Função Valor das Ações $Q^\pi(s, a) : (S \times A) \rightarrow \mathbb{R}$, que representa a soma das recompensas que serão obtidas considerando que: a) vigora o estado s ; b) o agente executou uma ação a ; c) deste ponto em diante, foi seguida uma política de ações π . O valor da ação é a

recompensa da ação mais o valor do estado que passa a vigorar a partir da ação do agente.

$$Q^\pi(s, a) = r(s, a) + \gamma V^\pi(\delta(s, a)) \quad (4.4.6)$$

Considerando o que foi apresentado até agora, afirma-se que a finalidade da aprendizagem por reforço é fazer com que o agente aprenda uma política de ações π^* ótima, que maximiza a função $V^\pi(V^*)$ ou a função $Q^\pi(Q^*)$, dado que:

$$\pi^* = \operatorname{argmax}_\pi [V^\pi(s)] \quad (4.4.7)$$

Em suma, a idéia é fazer com que o agente aprenda a agir de forma a maximizar suas recompensas futuras. Neste contexto, em um ambiente determinístico, e conhecidas as funções recompensa e de transição de estados, é possível computar, através de programação dinâmica (em tempo polinomial), uma política ótima, com:

$$V^*(s) = \max_a [r(s, a) + \gamma V^*(f(s, a))] \quad (4.4.8)$$

$$\pi^*(s) = \operatorname{argmax}_a [r(s, a) + \gamma V^*(f(s, a))] \quad (4.4.9)$$

Para ambientes não determinísticos, basta substituir a função de transição de estados $f(s, a)$ pelo modelo de transição de estados $\delta(s, a)$ (supostamente conhecido). Por fim, infere-se que computar todas essas funções, para ambientes muito complexos, pode implicar elevada complexidade de tempo, decorrendo problemas de escalabilidade. A solução para o processo de aprendizagem, dadas tais condições, pode residir na utilização de algoritmos simplificados com atuação independente da função de transição (*model-free*).

Desta feita, quando um agente não conhece a função ou o modelo de transição de estados, ou o ambiente é demasiado complexo, tornando impossível computar soluções adequadas, a programação dinâmica não pode ser utilizada. Contudo, é possível determinar π^* se Q^* é conhecido, sem a necessidade de conhecer f , δ ou r , com:

$$\pi^* = \operatorname{argmax}_a [Q(s, a)] \quad (4.4.10)$$

Isto posto, verifica-se que é possível estabelecer um processo de aprendizagem que não considera $V(s)$, substituída por $Q(s, a)$. Para aplicação no processo de aprendizagem, portanto, temos:

$$V^*(s) = \max_a [Q^*(s, a)] \quad (4.4.11)$$

$$\pi^*(s) = \operatorname{argmax}_a [Q^*(s, a)] \quad (4.4.12)$$

No processo em tela, verifica-se que os valores Q^* são definidos a partir das experiências do agente, em substituição a função de transição. Considerando o fator tempo na Função Valor das Ações, observa-se as expressões a seguir:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma V^*(\delta(s_t, a_t)) \quad (4.4.13)$$

De onde decorre:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \max_a [Q(s_{t+1}, a_{t+1})] \quad (4.4.14)$$

4.4.1 Algoritmo Q-learning

No contexto da Aprendizagem por Reforço existem diversos Algoritmos que poderiam ser aplicados no desenvolvimento do presente trabalho, contudo, segundo Kaelbling *et al* (1996) e Lin (1992), o Q-learning parece ser o algoritmo *model-free* mais eficaz para aprender mediante recompensas atrasadas. E tal condição é primordial para a aplicação proposta neste trabalho, dada a complexidade da função de transição de estados.

O Q-learning é um algoritmo de aprendizagem por reforço baseado em recompensa atrasada. O objetivo do algoritmo é, em cada intervalo de tempo, maximizar $Q(s, a)$, definida como a recompensa esperada por adotar a ação a no momento em que vigora um estado s . Para tanto, o algoritmo mantém e atualiza uma tabela de valores Q (um valor para cada combinação de estado-ação). A utilidade E de qualquer estado se define pelo máximo valor Q obtido considerando-se todas as ações que podem ser tomadas nesse estado. O valor Q de praticar uma ação (em um estado) é definido como a soma da recompensa imediata r com a utilidade $E(s')$ do próximo estado s' , definido segundo uma função de transição de estado T , descontado o fator γ , o qual será objeto de maiores comentários posteriormente. Considerando a aplicação da equação anterior, tem-se:

$$s' \leftarrow T(s, a) \quad (4.4.15)$$

$$E(s) = \max_a Q(s, a) \quad (4.4.16)$$

$$Q(s, a) = r + \gamma E(s'), 0 \leq \gamma \leq 1 \quad (4.4.17)$$

Uma vez estabelecida a Função Valor das Ações a ser empregada, decorre a necessidade de se definir a forma como as políticas serão aprendidas. Como a função Q torna a ação do agente explícita, os valores de Q podem ser estimados *on-line* mediante a utilização da regra de atualização do algoritmo $TD(0)$ (método de aprendizagem por diferença temporal) de Sutton (1988), apresentada a seguir.

$$V(s_t) \leftarrow V(s_t) + (r_{t+1} + \gamma V(s_{t+1}) - V(s_t)) \quad (4.4.18)$$

Conforme mostrado na expressão acima, sempre que um estado s_t é visitado, seu valor estimado é atualizado para estar mais perto de $r_{t+1} + \gamma V(s_{t+1})$, já que r_{t+1} é a recompensa instantânea recebida e $V(s_t)$ é o valor estimado do estado que de fato ocorre em seguida. A idéia fundamental é que $r_{t+1} + \gamma V(s_{t+1})$ representa uma amostra do valor de $V(s)$, e é mais provável que este valor seja correto, pois ele incorpora o verdadeiro valor da recompensa.

Neste ponto, cabe ressaltar que os valores de Q também devem ser usados para definir a política, posto que a ação pode ser definida escolhendo-se aquela com o máximo valor de Q , considerando o cenário atual. Define-se, então, a regra básica do algoritmo Q-learning (Russel e Norvig, 2003):

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (4.4.19)$$

Onde α representa uma taxa de aprendizagem que estabelece o ritmo de aprendizagem, definindo a sensibilidade do agente à recompensa dada a ele; e γ representa um fator de desconto aplicado ao Valor Q da melhor ação que o agente pode adotar na próxima etapa. De forma mais detalhada, pode-se dizer que taxa de aprendizagem α determina em que medida as informações recém adquiridas irão substituir as informações antigas. Uma taxa 0 faz com que o agente não aprenda, enquanto uma taxa 1 faz o agente considerar apenas as informações mais recentes.

Por sua vez, o fator de desconto temporal γ determina a importância das recompensas futuras; um fator de 0 faz com que o agente atue de forma oportunista, por considerar apenas recompensas atuais, ao passo que fatores próximos a 1 fazem com que o agente supervalorize altos valores de recompensa obtidos no longo prazo. Os valores Q divergem quando o fator de desconto é igual ou maior do que 1.

A regra acima, então, é aplicada na construção do algoritmo Q-learning, apresentado a seguir.

Algorithm 4.1 Algoritmo Q-learning

```
1 inicializar o valor de  $Q(s, a)$  arbitrariamente
2 for all episódios do
3   observar o estado inicial  $s$ 
4   repeat
5     escolher uma ação  $a$  para o estado  $s$ 
6     executar a ação  $a$ 
7     observar a recompensa  $r$  e o novo estado  $s_{t+1}$ 
8      $Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha[r(s_t, a_t) + \gamma \max_a Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)]$ 
9      $s \leftarrow s_{t+1}$ 
10  until  $s$  ser um estado final
11 end for
```

Considerando o algoritmo acima, dado que , para um determinado estado atual, o agente possui várias ações possíveis, percebe-se que a melhor ação a ser tomada vai depender do valor do histórico das recompensas atribuídas em função de estados/ações específicos. Dessa forma, quanto maior o valor da recompensa atribuída para aquele estado/ação, maiores são as chances de o agente escolher aquela ação novamente.

Neste ponto, cabe ressaltar que um dos grandes desafios para a aplicação da AR reside na busca pelo balanceamento adequado entre duas possibilidades: explorar novas ações ou utilizar ações para as quais o sistema já recebeu uma recompensa satisfatória. Ou seja, para obter a melhor recompensa possível, o sistema deve privilegiar ações que já foram testadas no passado e efetivamente produziram boas recompensas. Contudo, para saber que ações são essas, o sistema, em algum momento, precisa testar ações inéditas. Ante a este dilema, decorrem duas certezas, qual sejam, a de que a falha é certa quando há total exclusão de uma das duas possibilidades e que o segredo do sucesso reside no adequado balanceamento entre elas (Russel e Norvig, 2003).

Com relação a esta situação, já foram desenvolvidos sofisticados métodos matemáticos na tentativa de buscar o adequado balanceamento entre *exploration* / *exploitation*. Contudo, os referidos métodos fundamentam-se em elevado conhecimento sobre o ambiente de aplicação, o que se define a partir de fortes hipóteses sobre a distribuição de probabilidade dos valores obtidos como reforço. Tal condição invalida a aplicação em ambientes desconhecidos e motivou a adoção de uma política randômica de *exploration* / *exploitation* no presente trabalho.

Capítulo 5

MODELAGEM E IMPLEMENTAÇÃO

O número de variáveis e fatores envolvidos no Gerenciamento do Fluxo de Tráfego Aéreo, com destaque para a imprescindível análise dos impactos projetados em função das políticas de ajuste de fluxo idealizadas, torna impensável a tomada de decisão sem uma ferramenta computacional de suporte.

Contudo, o processo ATFM constitui um domínio cuja complexidade inibe a adoção de métodos e/ou ferramentas completamente automatizadas, mormente quando considerada a sua intrínseca relação com o controle de tráfego aéreo, caracterizado como uma atividade crítica e realizada em tempo real (Stoltz e Ky, 2001; Stoltz e Guerreau, 2002).

Assim sendo, qualquer metodologia aplicada na solução (viável) desse dilema deve apoiar a concepção de um sistema onde existam pelo menos dois agentes, a saber: o agente não humano (ferramenta computacional) que processa as variáveis/fatores e sugere ações (políticas/medidas de ajuste de fluxo) em função de determinado cenário, e o agente humano que tem o arbítrio de aceitar ou não tais sugestões.

Neste contexto, cabe destacar duas inferências principais. A primeira diz respeito à necessidade de se desenvolver um sistema computacional que possa “decidir” a partir de cenários (consideradas as principais variáveis e fatores) e, subseqüentemente, continuar decidindo considerando os impactos dessa decisão. A segunda está relacionada ao fato de que uma ferramenta computacional ATFM, que pode ser classificada como um sistema de apoio à decisão, terá sua eficiência fortemente condicionada pela forma como é modelada a interação homem-sistema.

De maneira mais específica, pode-se dizer que tal sistema será tão mais eficiente quanto maior for a sua capacidade de atender às expectativas do agente humano, considerando critérios por ele estabelecidos e a sua experiência, em um cenário onde o referido agente humano é o especialista em ATFM e o usuário do sistema.

5.1 O SISCONFLUX

Atualmente no Brasil, o loop decisório ATFM, no nível tático, é completamente focado no agente humano. O Gerente de Fluxo, para tomar sua decisão acerca da medida de controle a ser aplicada, dispõe apenas de ferramentas computacionais informativas, conforme discutido no Capítulo 2. Com base nestas informações, o elemento humano, empregando apenas sua experiência e habilidade, define, por exemplo, os atrasos nas decolagens de determinado aeroporto, de forma que a condição de desbalanceamento seja contornada.

Para condições de desbalanceamento geograficamente restritas e temporalmente exíguas, a tomada de decisão feita desta forma pode incorporar algum grau de eficácia, o que será definido pela experiência e habilidade do gerente de fluxo. Contudo, para condições de desbalanceamento com impacto prolongado sobre grandes áreas (vários aeroportos), a medida padronizada pelo CGNA é a suspensão das decolagens. Considerando a impossibilidade do agente humano processar todas as variáveis envolvidas, definir medidas de controle personalizadas (para cada aeroporto) e projetar o impacto das medidas (cenários futuros), esta realmente parece ser a medida mais segura.

Com a implementação do SISCONFLUX, o processo decisório ATFM tático sofrerá uma significativa alteração, na medida em que um novo agente atuará nas fases do referido processo, conforme pode ser observado na Figura 5.1.1.

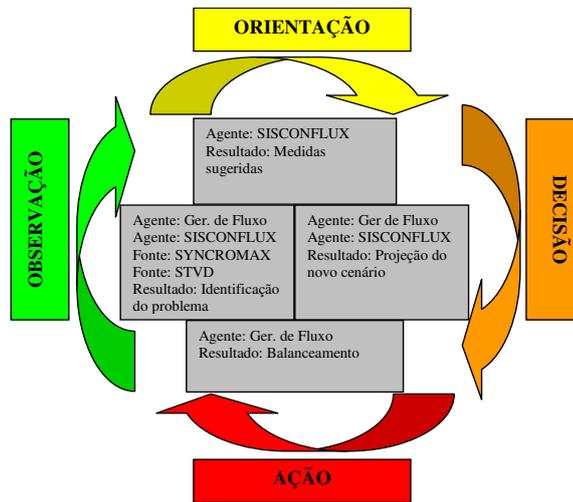


Figura 5.1.1: Processo Decisório ATFM com o SISCONFLUX

O SISCONFLUX é um sistema modular capaz de analisar o cenário aéreo e sugerir medidas restritivas de controle de fluxo de tráfego, de forma que os setores de controle não saturem. Em um primeiro momento, as medidas são geradas por algoritmos de fluxo máximo (MBF), e representam a solução mais adequada considerando apenas (como critério) a manutenção do melhor fluxo de tráfego possível observando-se a condição de não saturação dos setores de controle. A partir de determinado momento, as medidas sugeridas são geradas pelo algoritmo Q-learning (MAAD).

A Figura 5.1.2 nos mostra os módulos do SISCONFLUX e sua proposta de integração com os sistemas STVD e SYNCROMAX.

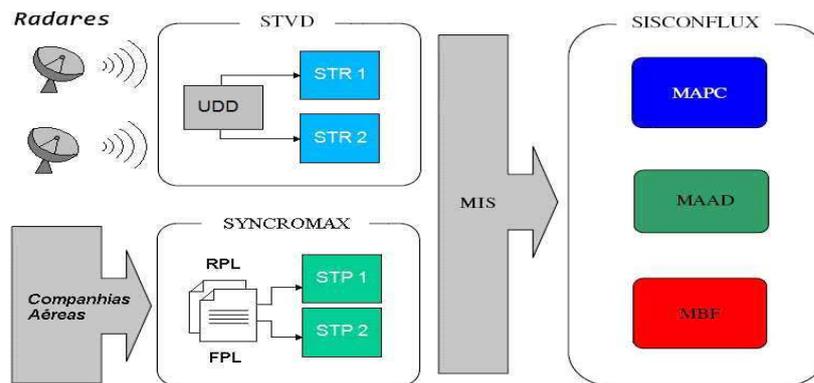


Figura 5.1.2: Integração do SISCONFLUX com o STVD e SYNCROMAX

A arquitetura do sistema prevê uma conformação modular, mediante o desenvolvimento de módulos específicos e interfaces para a integração com sistemas pré-existentes, conforme discutido a seguir. Nesta seção não constam comentários sobre o MAAD, posto que o mesmo é objeto de uma seção específica.

5.1.1 Interfaces com STVD e SYNCROMAX

O STVD é um sistema de Controle de Tráfego Aéreo (ATC) que processa as informações obtidas a partir dos radares (através de Unidades de Derivação Digital - UDD) que compõem o Sistema de Defesa Aérea e Controle de Tráfego Aéreo (SISDACTA) e apresenta, aos controladores de tráfego, o cenário referente ao tráfego aéreo em evolução na FIR-BS. O referido sistema processa ainda todas as informações dos movimentos aéreos com previsão de evolução nos espaços aéreos controlados, o que se dá através da compilação dos planos de voo (RPL e FPL) disponíveis.

Atualmente, o SYNCROMAX é um sistema voltado para o planejamento e o Gerenciamento de Fluxo de Tráfego Aéreo (ATFM) em caráter estratégico e pré-tático, disponibilizando um conjunto de indicadores e ferramentas utilizados para evitar o desbalanceamento entre a capacidade da infra-estrutura ATC e a demanda de fluxo de tráfego. A interface com os sistemas supracitados possibilita a extração das informações necessárias à composição do cenário, procedida pelo Módulo de Acompanhamento e Previsão de Cenário.

No desenvolvimento não foram utilizadas as informações disponibilizadas pelo Sistema de Tratamento Radar do STVD. O protótipo desenvolvido utilizou apenas as informações geradas a partir das bases de dados do SYNCROMAX (RPL) e do Sistema de Tratamento de Plano de Voo do STVD (FPL).

5.1.2 Módulo de Acompanhamento e Previsão de Cenário - MAPC

O Módulo de Acompanhamento e Previsão de Cenário (MAPC), parte integrante do SISCONFLUX, tem como uma de suas finalidades estimar a configuração futura da ocupação do espaço aéreo, constituindo um dispositivo computacional que realiza a previsão de cenário a partir das informações dos movimentos aéreos em evolução na Região de Informação de Voo de Brasília (FIR-BS), assim como as informações dos planos de voo (RPL e FPL). A Arquitetura do MAPC é apresentada na Figura 5.1.3.

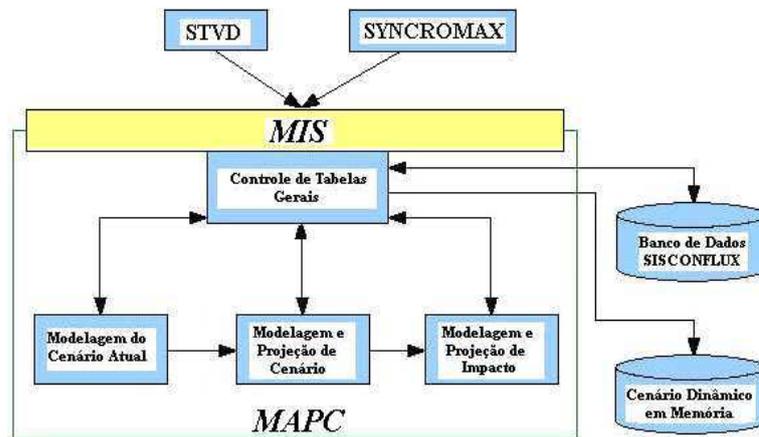


Figura 5.1.3: Arquitetura do MAPC (Sousa *et al*, 2008)

Contudo, a principal finalidade do módulo é projetar um novo cenário considerando a política de balanceamento sugerida pelo MBF e/ou MAAD, de forma que seja possível obter a nova configuração do espaço aéreo, em termos de quantidade de movimentos aéreos por setor de controle.

5.1.3 Módulo de Balanceamento de Fluxo (MBF)

O MBF (Sousa *et al*, 2008) opera em constante interação com o MAPC, além de ter atuação na conformação do conjunto de ações do MAAD. A arquitetura do MBF é apresentada na Figura 5.1.4.

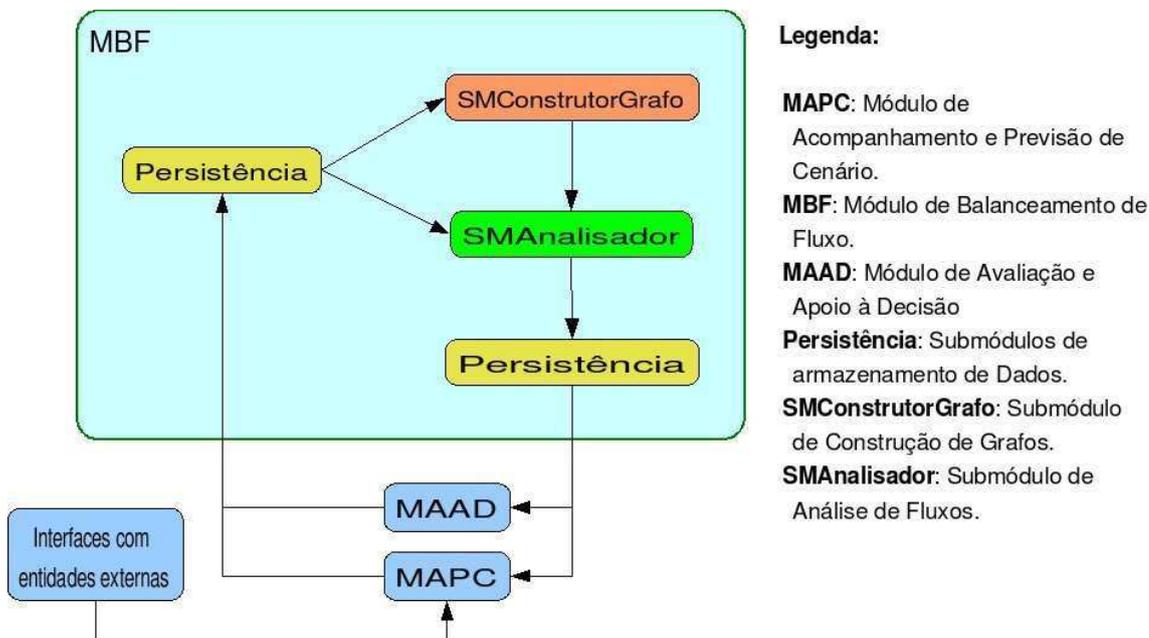


Figura 5.1.4: Arquitetura do MBF (Sousa *et al*, 2008)

Uma vez projetado o cenário concernente ao fluxo de tráfego no espaço aéreo controlado, o MBF realiza a análise da distribuição dos movimentos aéreos em

evolução, assim como as intenções de evolução (RPL e FPL), nos diversos setores de controle.

Ao serem verificados valores de saturação (ocupação igual ou superior 100% da capacidade do setor), este módulo inicia o processo de balanceamento do fluxo de tráfego, o que implica a projeção de um novo cenário em função do reordenamento dos horários de decolagem dos movimentos aéreos que partem das TMA.

Este módulo busca as possibilidades que impliquem a condição ideal para o fluxo de tráfego. Tal condição se caracteriza pela manutenção do maior fluxo de tráfego possível, observando-se a condição de que os setores de controle não atinjam os níveis de saturação. Uma vez que as deliberações estejam definidas, o MBF informa ao MAAD as medidas de controle de fluxo sugeridas (políticas de ajuste), as quais são incorporadas ao seu conjunto de ações.

5.2 Modelagem do Problema

Considerando a proposta de desenvolvimento de um agente computacional ATFM, faz necessário definir, em termos gerais e sob o enfoque da Inteligência Artificial, o modelamento do problema (compilação-leitura do cenário aéreo / avaliação do cenário aéreo / escolha de uma política de ajuste de fluxo de tráfego / aprendizagem-incorporação da experiência humana), considerando todos os elementos que fazem parte do processo em tela.

Conforme comentado no Capítulo 2, o espaço aéreo brasileiro está dividido em 46 setores de controle e possui 39 Áreas Terminais de Aproximação (TMA). Considerando que cada setor pode comportar mais de uma dezena de aeronaves (variação típica entre 0 e 16 aeronaves), e que a cada terminal podem ser aplicados seis diferentes montantes de atraso (espera em solo de 5, 7, 10, 15, 20 e 25 minutos para as aeronaves decolando), os conjuntos de ações e de estados seriam muito grandes.

Como consequência, sob o ponto de vista da aplicação do Algoritmo Q-learning, tal condição resultaria em uma Q-table inviável de ser montada e processada em um tempo computacional compatível com a demanda dos Gerentes de Fluxo de Tráfego e com um custo de espaço suportável. Assim sendo, objetivando tornar o tempo de execução aceitável e reduzir a complexidade de espaço, o problema foi modelado conforme definido a seguir.

5.2.1 Definição e Características do Agente

Dada a problemática ATFM apresentada no Capítulo 2, a proposta consiste na implementação de um Agente ATFM. De forma objetiva, ao Agente caberá sugerir medidas restritivas de controle de fluxo de tráfego aplicadas às aeronaves decolando de determinadas TMA. Tais medidas se caracterizam por atrasos nas decolagens (espera em solo) de 5, 7, 10, 15, 20 e 25 minutos. O objetivo do Agente é evitar a saturação dos setores de controle de tráfego aéreo produzindo o menor montante de atraso possível. Desta feita, as medidas de desempenho são: quantidade de Aeronaves nos setores de controle e montante de atraso imputado às aeronaves afetadas pelas medidas restritivas.

Sob o enfoque da Inteligência Artificial, a solução proposta apresentada neste trabalho configura-se no desenvolvimento de um Agente Inteligente com aprendizagem, tal como definido por Russel e Norvig (2003). Considerando os principais componentes conceituais de um Agente de Aprendizado, o papel do elemento de aprendizagem é desempenhado pelo Submódulo de Aprendizagem, ao passo que o papel do Crítico é desempenhado pelo Submódulo de Tratamento Avaliação de Cenário, ambos descritos a posteriori. O papel de elemento de desempenho cabe ao sistema como um todo.

5.2.2 Natureza do Ambiente

O ambiente de atuação do Agente ATFM é o espaço aéreo brasileiro e a inerente estrutura de gerenciamento do tráfego aéreo, tal como descrito no Capítulo 2. De forma estrita, tal ambiente é parcialmente observável e estocástico, na medida em que o Agente não pode "perceber" todas as aeronaves em evolução no espaço aéreo (aeronaves abaixo do limite de detecção radar). Da mesma forma, podem surgir no cenário aeronaves que decolaram (de aeródromos não controlados) sem o conhecimento do mesmo. Em adição, o comportamento do tráfego aéreo não pode ser previsto com exatidão, posto que condições meteorológicas (vento, tempestades) e variantes técnico-operacionais podem afetar a dinâmica do fluxo.

Contudo, algumas condições permitiram modelar o problema considerando o ambiente totalmente observável e determinístico. Primeiramente, segundo as normas em vigor no Sistema de Controle do Espaço Aéreo Brasileiro, as aeronaves em evolução em determinadas porções do espaço aéreo, desprovidas de cobertura radar (principalmente aeronaves voando em baixas altitudes), não recebem o serviço de controle de tráfego aéreo e, portanto, não constituem objeto para o gerenciamento do fluxo de tráfego.

Por outro lado, a concepção do sistema prevê que o cenário seja "inspecionado" constantemente, de acordo com os ciclos de atualização do sistema. Como consequência, os cenários projetados de forma determinística e as medidas restritivas definidas segundo um modelo determinístico são atualizados. Desse modo, as medidas sugeridas serão continuamente atualizadas até que não mais exista congestionamento nos setores de controle.

5.2.3 Conjunto de Ações

Conforme discutido previamente, a legislação brasileira (DECEA, 2007) prevê seis diferentes tipos de medidas de controle de fluxo de tráfego aéreo, quais sejam: espera no solo, esperas em rota, redução de velocidade, rotas alternativas, vetorações para atrasos em rota e pouso/espera em aeródromos intermediários.

Neste trabalho optou-se pela utilização da espera em solo. E tal escolha é plenamente justificada pois, sob o ponto de vista da segurança de vôo, é altamente recomendável que os procedimentos de gerenciamento de fluxo de tráfego afetem as aeronaves ainda em solo, de forma que pilotos e controladores não sejam obrigados a realizar ações não planejadas e que possam ter implicações operacionais tais como: aumento de consumo de combustível, aumento da quantidade de aeronaves em determinado setor do espaço aéreo, aumento da carga de trabalho de pilotos e controladores, dentre outras.

Em adição, quaisquer medidas que aumentem o tempo de vôo das aeronaves (tais como esperas em rota, rotas alternativas e vetoração para atrasos) geram relevantes impactos ambientais, posto que as mesmas permanecem mais tempo queimando combustíveis fósseis.

Para a implementação e teste do algoritmo utilizado pelos dois protótipos do MAAD, foram definidos dois conjuntos de ações distintos. O primeiro conjunto de ações, utilizado pelo primeiro protótipo, foi definido pelo SISCONFLUX de forma autônoma, a partir das ações de ajuste de fluxo produzidas pelo Módulo de Balanceamento de Fluxo (MBF). O segundo conjunto de ações, empregado no segundo protótipo, é formado pelo conjunto de ações inicial acrescido de ações efetivamente aplicadas pelos gerentes de fluxo do CGNA, em cenários reais, no período de 2008 a 2010. Tais políticas foram extraídas dos Relatórios de Situação Diária de gerenciamento de fluxo de tráfego do CGNA.

Em tais relatórios encontra-se uma análise diária pormenorizada da condição de operação no espaço aéreo brasileiro, incluindo todas as medidas de controle

de fluxo de tráfego aplicadas no país, bem como a descrição dos fatos (condições técnicas e operacionais de auxílios à navegação, pistas de pouso e pátios de manobras, fenômenos meteorológicos, etc.) que as geraram. Dessa forma, o sistema funcionou com duas configurações distintas, quais sejam: atuação autônoma e atuação mediante incorporação da experiência do agente humano. Neste ponto reside a principal contribuição do trabalho, na medida em que foi possível estabelecer uma análise comparativa do desempenho do sistema nas condições supracitadas.

Estes conjuntos de ações doravante serão referenciados como Medidas de Ajuste de Fluxo (ou simplesmente Medidas). Para a definição das Medidas, foram pesquisados todos registros (Relatórios de Situação Diária) feitos entre janeiro de 2008 e março de 2010. De forma geral, uma Medida é constituída por atrasos (esperas em solo) de 5, 7, 10, 15, 20 e 25 minutos imputados aos horários de decolagem das aeronaves partindo de determinada(s) TMA.

Esses atrasos podem ser aplicados indistintamente a todas as aeronaves partindo de uma TMA ou apenas em aeronaves com destinos específicos. Dentre as 39 TMA existentes, foram consideradas apenas aquelas com movimento anual superior a 20.000 pousos/decolagens, totalizando 19 TMA e aproximadamente 93% do movimento nacional computado pela INFRAERO (INFRAERO, 2009). A tabela 5.2.1 apresenta dois exemplos de ações (Medidas) pertencentes ao conjunto de ações dos protótipos desenvolvidos.

Tabela 5.2.1: Medidas de Ajuste de Fluxo

<i>Medida 1</i>						
<i>TMA</i>	5	7	10	15	20	25
<i>SP</i>			X			
<i>BR</i>					X	
<i>BH</i>			X			
<i>AN</i>			X			
<i>CY</i>	-	-	-	-	-	-
<i>UL</i>			X			
<i>RJ</i>		X				
<i>CT</i>				X		
<i>PA</i>					X	
<i>VT</i>	-	-	-	-	-	-
<i>FL</i>		X				
<i>CG</i>				X		
<i>LO</i>			X			
<i>SV</i>		X				
<i>RF</i>		X				
<i>FZ</i>		X				
<i>PS</i>	-	-	-	-	-	-
<i>NT</i>	-	-	-	-	-	-
<i>MN</i>		X				
<i>BE</i>		X				
<i>ST</i>	-	-	-	-	-	-

<i>Medida 2</i>						
<i>TMA</i>	5	7	10	15	20	25
<i>SP</i>			RE			
<i>BR</i>			SP			
<i>BH</i>			SP			
<i>AN</i>	-	-	-	-	-	-
<i>CY</i>	-	-	-	-	-	-
<i>UL</i>	-	-	-	-	-	-
<i>RJ</i>	-	-	-	-	-	-
<i>CT</i>		X				
<i>PA</i>	X					
<i>VT</i>	-	-	-	-	-	-
<i>FL</i>	-	-	-	-	-	-
<i>CG</i>	X					
<i>LO</i>	-	-	-	-	-	-
<i>SV</i>	-	-	-	-	-	-
<i>RF</i>	-	-	-	-	-	-
<i>FZ</i>	-	-	-	-	-	-
<i>PS</i>	-	-	-	-	-	-
<i>NT</i>	-	-	-	-	-	-
<i>MN</i>	X					
<i>BE</i>	X					
<i>ST</i>	-	-	-	-	-	-

Conforme pode ser observado na tabela acima, uma ação é constituída por um conjunto de atrasos especificados para cada terminal. Quando o atraso deve ser aplicado a todas as aeronaves partindo de determinada TMA utiliza-se o simbolo "X" para definir tal condição. Por outro lado, quando o atraso deve ser aplicado apenas para aeronaves com um destino determinado, utiliza-se as duas letras que representam a TMA de destino, por exemplo "SP", ou a FIR de destino, por exemplo "RE".

5.2.4 Conjunto de Estados

O Algoritmo Q-Learning utiliza uma tabela denominada Q-Table a qual, dentro do processo de aprendizagem, incorpora a experiência do agente; ou seja, ela armazena conjuntos Ação-Cenário que representam os estados observados pelo sistema. A configuração do cenário se dá com relação à quantidade de aeronaves nos setores de controle em determinado horário. A distribuição típica de aeronaves nos setores de controle de tráfego aéreo varia de 0 a 16; estes valores, contudo, podem chegar a mais de 20.

Portanto, colocada uma hipótese otimista, o conjunto de estados possíveis, considerando a totalidade do espaço aéreo brasileiro, alcança uma grandeza

da ordem de 46^{17} . Na Tabela 5.2.2 apresenta-se a distribuição de aeronaves no espaço aéreo da FIR-BS, na data de 12/03/2010, em horários escolhidos aleatoriamente.

Tabela 5.2.2: Cenário Aéreo Projetado para a FIR BS em 12/03/2010

Hora	Setor 1-2	Setor 3-4	Setor 5	Setor 6	Setor 7	Setor 8	Setor 9	Setor 10	Setor 11	Setor 12
00:15	19	5	3	2	2	1	7	7	10	5
02:15	10	2	4	8	9	1	5	9	4	3
10:30	13	8	6	14	10	1	5	9	3	4
12:40	10	10	10	9	1	4	4	4	5	2
14:10	17	7	5	12	9	2	5	10	9	3
15:25	10	13	9	7	2	2	2	2	3	4
21:00	10	6	7	7	7	8	1	11	10	9
23:00	9	11	16	6	4	0	8	6	13	7
23:25	11	14	10	7	5	2	4	12	9	3

Um conjunto de estados com essa magnitude implica elevada complexidade de tempo e de espaço, o que dificulta a implementação de um processo de aprendizagem baseado em tabelas (Q-Table). Como consequência, para efeito de composição da Q-table, foram definidas duas condições: Setor Saturado (quantidade de aeronaves igual ou superior a 14 para setores desagrupados e 12 para setores agrupados) e Setor Normal (quantidade de aeronaves menor do que 14 para setores desagrupados e 12 para setores agrupados).

Com a simplificação sugerida, o conjunto de estados inerente ao cenário aéreo brasileiro tem uma dimensão máxima igual a 46^2 . A Tabela 5.2.3 ilustra a simplificação sugerida, considerado o cenário referente à Tabela 5.2.2, onde "S" indica setor saturado e "N" indica setor não saturado.

Tabela 5.2.3: Cenário Aéreo Projetado para a FIR BS em 12/03/2010 - Compilação Simplificada

Hora	Setor 1-2	Setor 3-4	Setor 5	Setor 6	Setor 7	Setor 8	Setor 9	Setor 10	Setor 11	Setor 12
00:15	S	N	N	N	N	N	N	N	N	N
02:15	N	N	N	N	N	N	N	N	N	N
10:30	S	N	N	S	N	N	N	N	N	N
12:40	N	N	N	N	N	N	N	N	N	N
14:10	S	N	N	N	N	N	N	N	N	N
15:25	N	S	N	N	N	N	N	N	N	N
21:00	N	N	N	N	N	N	N	N	N	N
23:00	N	N	S	N	N	N	N	N	N	N
23:25	N	S	N	N	N	N	N	N	N	N

5.3 Concepção e Arquitetura do MAAD

O Módulo de Avaliação e Apoio à Decisão tem como principal funcionalidade gerar medidas de ajuste de fluxo de tráfego em função de determinado cenário. Para tanto, o MAAD conta com três submódulos, conforme pode ser observado na Figura 5.3.1, os quais serão descritos a seguir.

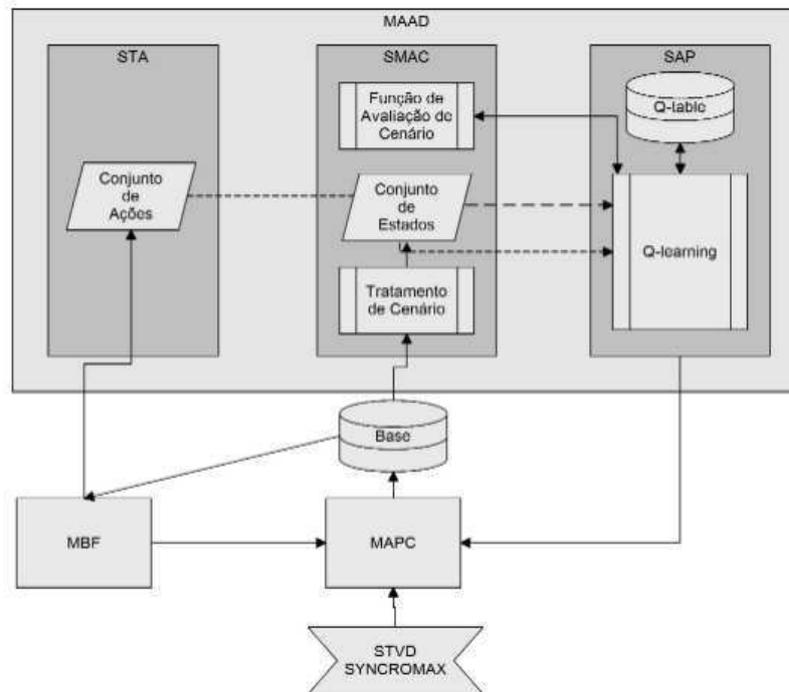


Figura 5.3.1: Arquitetura do MAAD

Conforme pode ser observado na Figura 3.5.1, o fluxo de dados e informações no sistema ocorre da seguinte forma:

- os arquivos inerentes às demandas de tráfego são geradas pelo SYNCROMAX/STVD e são direcionadas para o MAPC;
- o MAPC projeta as demandas, gerando tabelas de ocupação dos setores de controle, que são armazenadas na Base de Dados do sistema;
- o MBF busca as tabelas de ocupação dos setores de controle, converte as demandas de tráfego em multi-fluxos (Teoria de Grafos) e, caso uma condição de saturação seja verificada, gera ações (medidas de controle de fluxo de tráfego); as ações são agregadas ao Conjunto de Ações do MAAD, através do Submódulo de Tratamento de Ações (STA);
- o MAPC recebe as ações sugeridas, projeta um novo cenário considerando as medidas geradas pelo MBF e atualiza a base do sistema;
- o Submódulo de Avaliação de Cenário (SMAC), a partir das saídas do MAPC, faz o mapeamento do cenário aéreo, produzindo a tabela de es-

tado (Tratamento de Cenário), tal como discutido em 5.2.4, em função das condições não saturado e saturado;

- as tabelas de estado produzidas passam a compor o Conjunto de Estados do MAAD;
- O Algoritmo Q-learning, implementado no Submódulo de Aprendizagem (SAP), para um determinado estado em vigor, busca uma ação (nova) no Conjunto de Ações ou busca na Q-table, caso já exista, uma ação já adotada anteriormente para aquele estado.
- a partir da definição, pelo Q-learning, da ação (a) para o estado (s), o SMAC (Função de Avaliação de Cenário) calcula o valor de $Q(s, a)$, que é retornado ao Q-learning para o cálculo da recompensa do agente e para a atualização da Q-table;
- o SAP alimenta o MAPC com a ação (a) definida pelo Q-learning, ocorrendo a projeção de um novo cenário e o reinício do ciclo.

5.4 Função de Avaliação de Cenário

No desenvolvimento da estrutura de recompensa, para um agente baseado em Aprendizagem por Reforço, é de fundamental importância a definição de uma função de avaliação de estado (cenário) que efetivamente reflita os objetivos atribuídos ao referido agente, ou seja, a otimização (minimização ou maximização) de determinados parâmetros.

No presente trabalho, foram desenvolvidas duas funções de avaliação de cenário. Estas funções incorporam critérios e parâmetros definidos pelo CGNA. Tais critérios são: montante de atraso em TMA, quantidade de aeronaves nos setores ATC e complexidade dos setores ATC. Derivam-se portanto as funções de avaliação, onde φ , β e χ são constantes normalizadoras.

$$T_C(e) = -(\varphi D(e) + \beta C(e)), \varphi + \beta = 1 \quad (5.4.1)$$

$$T_N(e) = -(\varphi D(e) + \chi N(e)), \varphi + \chi = 1 \quad (5.4.2)$$

No presente trabalho, apenas a primeira função foi utilizada, ou seja, os experimentos realizados consideraram os fatores montante de atraso em TMA ($D(e)$) e quantidade de aeronaves nos setores ATC ($C(e)$). Assim sendo, as constantes φ e β foram dimensionadas de forma que, no pior cenário possível, a ponderação de cada fator fosse de aproximadamente 50%; o pior cenário foi definido da seguinte forma:

- Cada aeroporto (nas TMA) apresentando uma demanda de 1 decolagem a cada 3 minutos, perfazendo um total de 40 decolagens em 120 minutos (por aeroporto).
- Para cada TMA considera-se a aplicação de uma espera em solo de 25 minutos (valor máximo).
- Todos os setores de controle saturados.

Em suma, o pior cenário se define por todos os setores saturados e todas as TMA recebendo a maior restrição de tráfego possível (25 minutos), considerando-se que, em cada aeroporto das TMA, 40 aeronaves observarão (projeção de 120 minutos) o atraso sugerido. Na subseção 5.4.5 são apresentados os valores de $D(e)$, $C(e)$ e $T(e)$ para o pior caso idealizado.

5.4.1 Fator Montante de Atraso - $D(e)$

Este critério tem por finalidade avaliar o cenário em função do montante de atraso produzido, nas TMA impactadas, pela política de balanceamento. O montante de atraso será estabelecido para cada uma das TMA, que podem conter um ou mais aeroportos, e será o produto do número de movimentos aéreos previstos para decolar (nos próximos 120 minutos) pela política sugerida (tempo de atraso em minutos), conforme definido a seguir.

$$D(e) = \sum_{t \in T} D_t(e) \quad (5.4.3)$$

Com:

$$D_t(e) = m_t \cdot p_t \quad (5.4.4)$$

Onde:

- e - Estado do cenário;
- t - Terminal (TMA) impactada por medida de controle de fluxo;
- T - Conjunto de terminais (TMA) impactadas no cenário;
- m_t - Quantidade de aeronaves impactadas na terminal t a cada ciclo de atualização do sistema;
- p_t - Medida restritiva (atraso em minutos) aplicada à terminal (TMA) t .

5.4.2 Fator Quantidade de Aeronaves nos Setores ATC - $C(e)$

A quantidade de aeronaves nos setores ATC, apesar de não ser o fator preponderante na determinação da carga de trabalho imposta aos controladores de

tráfego aéreo como decorrência de uma medida ATFM, constitui um critério a ser utilizado na função de avaliação de cenário, de forma que a mesma possa ser utilizada na impossibilidade de se calcular a Complexidade ATC. No contexto atual, apenas alguns setores do espaço aéreo brasileiro estão mapeados quanto à Complexidade ATC, motivo pelo qual este critério foi usado no estudo de caso apresentado neste trabalho.

$$C(e) = \sum_{s \in S} C_s(e) \quad (5.4.5)$$

Com:

$$C_s(e) = 10, \text{ para } (m_s - c_s) \geq 0 \quad (5.4.6)$$

$$C_s(e) = 0, \text{ para } (m_s - c_s) < 0 \quad (5.4.7)$$

Onde:

s - Setor ATC;

S - Conjunto de Setores ATC existentes no cenário;

m_s - Quantidade de aeronaves no setor ATC s ;

c_s - Capacidade do Setor ATC s .

Os valores constantes apresentados nas equações acima foram obtidos a partir da equação $C_s(e) = \Theta(m_s - c_s)e^{(m_s - c_s)}$, adaptada de Tumer e Agogino (2007). Através da aplicação da equação nos cenários estudados, e segundo a distribuição de probabilidade dos resultados obtidos, foi definido o valor médio inerente a cenários saturados (10) e cenários não saturados (0).

5.4.3 Fator Densidade Dinâmica

A complexidade total de um setor de controle, ou Densidade Dinâmica, é a soma ponderada da contribuição individual dos fatores de complexidade. Cada um destes fatores contribui para a complexidade total com um valor máximo (MAX), computado entre o tempo t e o tempo $t + \tau$, ou com uma soma (\sum) neste mesmo o intervalo de tempo. O Fator Complexidade ATC do cenário $N(e)$, apresentado a seguir, é o somatório da complexidade de cada um dos setores ATC. A sistemática de avaliação de complexidade apresentada neste trabalho foi desenvolvida pela NASA (NASA, 1996).

$$N(e) = \sum_{s \in S} N_s(e) \quad (5.4.8)$$

Com:

$$N_s(e) = \sum_1^{12} k_n \cdot \eta_n \quad (5.4.9)$$

Onde:

η - fator de complexidade;

k - fator de ponderação.

Os fatores de complexidade, bem como os fatores de ponderação a eles associados, estão listados no trabalho da NASA supracitado. Tais fatores referem-se às características do espaço aéreo (estruturação), características do fluxo de tráfego (dinâmica) e características das aeronaves em evolução. No presente trabalho não constam maiores detalhes sobre o cálculo da Densidade Dinâmica, posto que os estudos de caso analisados utilizam apenas o Fator Quantidade de Aeronaves.

5.4.4 Avaliação de Cenário

O MAAD foi concebido como um agente que tem como objetivo escolher a ação que implique na melhor avaliação possível do sistema (recompensa), a qual se dá através da função de avaliação expressa pela equação 5.4.1. A partir desta equação, o agente busca maximizar as recompensas mediante a utilização do algoritmo Q-Learning. Através da Tabela 5.4.1, pode-se verificar como se dá o cálculo de $D(e)$ a partir dos valores de $D_t(e)$ estabelecidos para cada área terminal de aproximação (TMA); nas Tabelas 5.4.2 e 5.4.3 verifica-se como ocorre o cálculo de $C(e)$, a partir dos valores de $C_s(e)$ estabelecidos para cada setor de controle.

A Tabela 5.4.1 apresenta um exemplo da aplicação da referida equação, inerente às condições do espaço aéreo no dia 18 de março de 2009, no horário de 10:15Hs. A título de ilustração, na TMA-SP existem 69 aeronaves com previsão de decolagem nas próximas duas horas (entre 10:15Hs - 12:15Hs), e o sistema sugeriu uma medida restritiva de 7 minutos para essa TMA; ou seja, cada aeronave sofrerá um atraso de 7 minutos no seu horário prévio de decolagem. Portanto, a TMA-SP contribui com um $D_t(e) = 483$ (valor não normalizado); por sua vez, a TMA-BH contribui com um $D_t(e) = 375$. Após o cômputo (somatório) de todas as TMA, tem-se um $D(e) = 23,739501$ (valor já normalizado por φ).

Tabela 5.4.1: Cálculo de $T(e)$ - Cenário Típico

09:55						
<i>Crítérios /</i>	m_t	p_i	D_t	$D(e)$	$C(e)$	$T(e)$
<i>TMA</i>						
TMA-SP	69	7	483	23,739501	19,72603	43,465531
TMA-BR	16	10	160			
TMA-BH	25	15	375			
TMA-AN	4	5	20			
TMA-UL	7	5	35			
TMA-RJ	37	7	259			
TMA-CT	18	7	126			
TMA-PA	12	10	120			
TMA-FL	4	10	40			
TMA-SV	7	5	35			
TMA-RE	6	5	30			
TMA-FZ	4	5	20			
TMA-MN	2	5	10			
TMA-BE	4	5	20			

Para exemplificar a obtenção dos valores inerentes a $C(e)$, deve-se observar as Tabelas 5.4.2 e 5.4.3. As referidas tabelas apresentam a quantidade de aeronaves nos setores de controle da FIR Brasília (FIR BS) e da FIR Curitiba (FIR CW), considerando a demanda prevista para o horário de 10:15Hs. Tais quantidades são convertidas conforme estabelecido pelas equações 5.4.6 e 5.4.7, obtendo-se o $C_s(e)$ para cada um dos setores de controle e o valor global $C(e)$ inerente às FIR BS e FIR CW.

Tabela 5.4.2: $C(e)$ inerente à FIR BS - 10:15Hs

		FIR BS								$C(e)$
		1.1-2	1.3-4	1.5	1.6	1.7-9	1.10	1.11	1.12	
<i>Qtd. Anv.</i>		13	10	6	5	11	8	4	2	
$C_s(e)$		10	0	0	0	0	0	0	0	10

Tabela 5.4.3: $C(e)$ inerente à FIR CW - 10:15Hs

		FIR CW					$C(e)$
		2.1-2	2.3	2.4-5	2.6-8	2.9-10	
<i>Qtd. Anv.</i>		20	13	9	2	2	
$C_s(e)$		10	0	0	0	0	10

Por exemplo, na Tabela 5.4.2, o setor 1.5 tem 6 aeronaves (valor inferior ao limite de saturação - 14), de modo que tal setor contribui com um $C_s(e) = 0$; o setor agrupado 1.1-2, por sua vez, tem 13 aeronaves (valor superior ao limite de saturação - 12), contribuindo com um $C_s(e) = 10$. Desta forma, a FIR BS contribui com um $C(e) = 10$.

Na Tabela 5.4.3, verifica-se que o setor agrupado 2.1-2 apresenta 20 aeronaves; como tal valor é superior ao limite de saturação, este setor contribui com

um $C_s(e) = 10$. No presente exemplo, a FIR Recife e a FIR Amazônica contribuíram com $C(e) = 0$. Portanto, o somatório de todos os $C_s(e)$ resulta no valor de $C(e) = 19,72603$ (valor já normalizado por β). Por fim, obtém-se o valor final de $T(e) = -43,465531$, que corresponde à avaliação do cenário segundo os critérios e parâmetros estabelecidos.

5.4.5 Avaliação de Cenário - Pior Caso

A partir dos valores não normalizados de $C(e)$ e $D(e)$, obtidos na condição de pior cenário, foram definidos os valores das constantes φ e β , de forma que a ponderação de cada um destes fatores contribuísse com aproximadamente 50% do valor final de $T(e)$. A Tabela 5.4.4 apresenta o cálculo de $T(e)$ na condição do pior cenário idealizado.

No pior cenário, a TMA-SP contribui com um $D_t(e) = 5.000$ (valor não normalizado). Após o cômputo de todas as TMA, tem-se um $D(e) = 246,573$ (valor já normalizado por φ).

Tabela 5.4.4: Cálculo de $T(e)$ - Pior Cenário

10:55						
<i>Critérios /</i>	<i>m_t</i>	<i>p_i</i>	<i>D_t</i>	<i>D(e)</i>	<i>C(e)</i>	<i>T(e)</i>
TMA						
<i>TMA-SP</i>	120	25	3000	246,573	246,57538	493,14838
<i>TMA-BR</i>	40	25	1000			
<i>TMA-BH</i>	80	25	2000			
<i>TMA-AN</i>	40	25	1000			
<i>TMA-CY</i>	40	25	1000			
<i>TMA-UL</i>	40	25	1000			
<i>TMA-RJ</i>	80	25	2000			
<i>TMA-CT</i>	40	25	1000			
<i>TMA-PA</i>	40	25	1000			
<i>TMA-VT</i>	40	25	1000			
<i>TMA-FL</i>	40	25	1000			
<i>TMA-CG</i>	40	25	1000			
<i>TMA-LO</i>	40	25	1000			
<i>TMA-SV</i>	40	25	1000			
<i>TMA-RE</i>	40	25	1000			
<i>TMA-FZ</i>	40	25	1000			
<i>TMA-PS</i>	40	25	1000			
<i>TMA-NT</i>	40	25	1000			
<i>TMA-MN</i>	40	25	1000			
<i>TMA-BE</i>	40	25	1000			
<i>TMA-ST</i>	40	25	1000			

Para a obtenção dos valores inerentes a $C(e)$, consideram-se todos os setores saturados e, mediante a aplicação das equações 5.3.5 e 5.3.6, tem-se $C_s(e) = 10$, para cada um dos 25 setores de controle, e o somatório de todos os $C_s(e)$ resulta no valor de $C(e) = 246,57538$ (valor já normalizado por β). Por fim, obtém-se o valor final de $T(e) = -493,14838$.

5.5 Nova Estrutura de Aprendizagem nas Aplicações ATFM

Conforme discutido no Capítulo 3, este trabalho foi precedido por outros estudos considerando a aplicação de agentes baseados em Aprendizagem por Reforço na solução da problemática ATFM, a partir dos quais foi possível verificar que a implementação e/ou a adaptação de um sistema de agentes para a gestão do tráfego aéreo demanda conhecimento intensivo (Wolf, 2007).

Alguns desses conhecimentos podem ser obtidos a partir de especialistas e em tempo real, mas nem sempre essa é uma possibilidade prática, ainda mais quando se verifica que determinado conhecimento não é totalmente dominado mesmo por especialistas. Considerando tal condição, a aquisição de conhecimento pode se dar de algumas formas, como, por exemplo, a partir de dados históricos. Estes dados históricos, por sua vez, podem ser constituídos por um conjunto de ações aplicadas pelos especialistas humanos em cenários passados; tal modalidade de aquisição de conhecimento se define como aprendizagem a partir de dados históricos (Wolf, 2007).

Na aprendizagem a partir de dados históricos, os dados históricos, quando disponíveis, podem servir como base para o desenvolvimento de modelos de comportamento humano, manualmente ou através de métodos de aprendizagem de máquina; as condições do ambiente/cenário e os *inputs* dos especialistas humanos devem formar o espaço de atributos, a partir do qual a decisão correta será aprendida.

Neste ponto cabe ressaltar que o levantamento realizado, acerca de pesquisas empregando agentes inteligentes na problemática ATFM, indica que os estudos realizados apenas simularam o comportamento dos controladores de tráfego aéreo/gerentes de fluxo de tráfego aéreo, decorrendo que os agentes não "aprendiam" considerando o comportamento real dos agentes humanos.

O trabalho de Agogino e Tumer (2009), a título de exemplo, contempla a forma usualmente utilizada no processo de aprendizagem para agentes inteligentes ATFM, conforme pode ser observado na Figura 5.5.1.

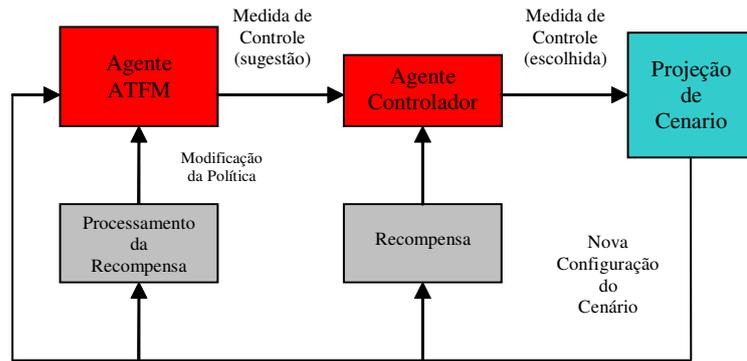


Figura 5.5.1: Sistema Semi-automatizado (Agogino e Tumer, 2009)

Na Figura 5.5.1, observa-se um modelo onde um Agente ATFM - gerador de ações de controle de fluxo de tráfego aéreo (medidas ATFM), baseado em aprendizagem por reforço, sugere medidas ATFM para um agente controlador de tráfego aéreo (Agente Controlador), ao qual cabe acatar ou não a sugestão fornecida. Contudo, no referido estudo, o Agente Controlador também é um agente computacional baseado em aprendizagem por reforço. Assim sendo, não é possível estabelecer uma análise consistente sobre o desempenho do sistema considerando a incorporação de aspectos comportamentais inerentes ao Agente Humano.

Desta feita, no presente trabalho propõe-se uma metodologia ainda não explorada em pesquisas na área da computação aplicadas a problemática ATFM. Nesta nova proposta, a experiência dos especialistas ATFM (humanos) é incorporada pelo Agente Computacional mediante a utilização de dados históricos, definidos a partir das medidas ATFM aplicadas pelos gerentes de fluxo do CGNA ao longo dos anos de 2008, 2009 e 2010. A Figura 5.5.2 apresenta, de forma esquemática, o ciclo de aprendizagem segundo a proposta em tela.

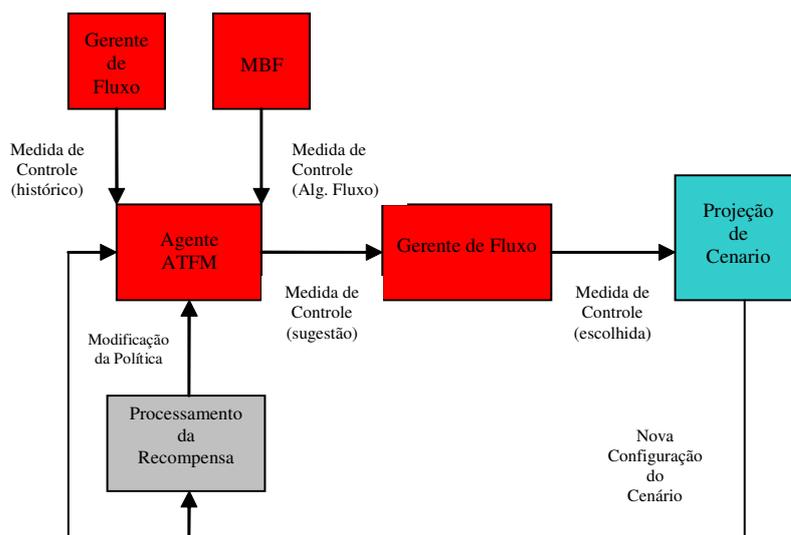


Figura 5.5.2: Proposta do MAAD

Conforme pode ser observado na Figura 5.5.2, o Agente ATFM (MAAD) é alimentado com ações (medidas ATFM) efetivamente utilizadas pelos Gerentes de Fluxo do CGNA em cenários reais (dados históricos); tais medidas passam a compor, juntamente com medidas geradas internamente pelo SISCONFLUX (MBF), o conjunto de ações do Agente. A partir do conjunto de ações construído, o Agente sugere ações e o ciclo de atuação/aprendizagem prossegue de forma típica.

Em síntese, a Figura 5.5.2 se diferencia da Figura 5.5.1 por representar uma concepção de aprendizagem onde se verifica a efetiva incorporação da experiência do especialista humano. Tal concepção agrega a este trabalho uma nítida vantagem em relação as pesquisas predecessoras, na medida em que permitiu a avaliação (análise comparativa) de ações típicas do agente humano e ações típicas do agente computacional.

Ressalta-se que, a partir da aplicação desta concepção, foi possível resolver um problema de extrema relevância no contexto da problemática ATFM, qual seja, a necessidade de se aplicar medidas de controle de fluxo de forma mais pontual e seletiva, atingindo aeronaves com destinos específicos. Esta aplicação tem por objetivo reduzir os impactos das referidas medidas a determinadas áreas do espaço aéreo, mais diretamente relacionadas ao desbalanceamento demanda x capacidade. Por fim, acrescenta-se que tais medidas não poderiam ser geradas por um sistema dotado apenas de algoritmos de fluxo máximo em rede, tais como os utilizados pelo MBF.

Capítulo 6

ESTUDO DE CASO

Para a realização do estudo de caso, foram utilizados dois protótipos que diferenciam-se exclusivamente pelo conjunto de ações. O primeiro protótipo, MAAD, agrega um conjunto de ações gerados exclusivamente pelo Módulo de Balanceamento de Fluxo (MBF). O segundo, MAAD* protótipo utiliza um conjunto de ações que incorpora um subconjunto formado pelas ações aplicadas pelos gerentes de fluxo de tráfego do CGNA.

A criação de um conjunto de ações que incorpora as ações efetivamente aplicadas pelos gerentes de fluxo do CGNA constitui umas das principais contribuições do presente trabalho. Conforme mencionado previamente, tal subconjunto foi montado a partir das informações contidas nos Relatórios de Situação Diária referentes ao período compreendido entre janeiro de 2008 e março de 2010. Através deste conjunto de ações, foi possível agregar ao protótipo MAAD* a experiência dos especialistas humanos, qual seja, o gerente de fluxo de tráfego aéreo do CGNA.

Os cenários submetidos aos protótipos foram definidos tomando como base as reais previsões de demanda de fluxo de tráfego disponibilizadas pelo CGNA. Consideradas as premissas básicas inerentes à agentes baseados em aprendizagem por reforço, os protótipos foram treinados mediante a utilização de cenários congestionados projetados a partir das previsões de demanda disponibilizadas. Para acelerar o processo de aprendizagem, foi adotada uma configuração de cenário, prevista na concepção operacional dos órgãos de controle e detalhada posteriormente, que reduz a capacidade do espaço aéreo e, conseqüentemente, aumenta os pontos de saturação.

6.1 Planejamento do Estudo de Caso

O estudo de caso ora apresentado foi desenvolvido através das seguintes fases:

1. Coleta de dados inerentes à demanda de tráfego aéreo (vôos) para o espaço aéreo brasileiro, considerando o período compreendido entre janeiro de 2008 e março de 2010.
2. Coleta de dados inerentes às medidas de controle de fluxo de tráfego aéreo aplicadas pelos gerentes de fluxo de tráfego do CGNA, em cenários reais, considerando o período compreendido entre janeiro de 2008 e março de 2010; tais medidas (ações) foram incorporadas ao conjunto de ações de um dos protótipos utilizados (protótipo MAAD*).
3. Inserção dos vôos no banco de dados do sistema, de modo que a ocupação dos setores de controle pudessem ser projetadas (quantidade de aeronaves em cada setor de controle).
4. Definição da configuração dos cenários, considerando os agrupamentos de setores de controle possíveis, segundo os modelos operacionais dos órgãos de controle de tráfego aéreo; as datas referentes aos cenários escolhidos foram definidas de forma que se pudesse explorar uma importante característica da circulação aérea, qual seja, o caráter repetitivo da demanda em função do dia da semana e da hora do dia.
5. Teste inicial de validação dos protótipos (Agente ATFM desprovido da experiência humana - MAAD / Agente ATFM com a experiência humana incorporada - MAAD*), considerando o cenário aéreo do dia 06 de março de 2008.
6. Verificação de convergência do algoritmo Q-learning empregado no protótipo MAAD, considerando o cenário aéreo do dia 18 de março de 2010.
7. Treinamento do protótipo MAAD, mediante o processamento de 20 cenários aéreos diferentes e aleatoriamente escolhidos.
8. Aplicação do protótipo MAAD considerando o cenário aéreo do dia 17 de março de 2010.
9. Verificação de convergência do algoritmo Q-learning empregado no protótipo MAAD*, considerando o cenário aéreo do dia 18 de março de 2010.
10. Treinamento do protótipo MAAD*, mediante o processamento de 20 cenários aéreos diferentes e aleatoriamente escolhidos.
11. Aplicação do protótipo MAAD* considerando o cenário aéreo do dia 17 de março de 2010.
12. Discussão dos resultados (análise comparativa do desempenho dos protótipos).

6.2 Configuração dos Cenários

No desenvolvimento do presente estudo, foram utilizados cenários aéreos projetados segundo a demanda real de tráfego aéreo observada em períodos específicos. Os protótipos processaram cenários em quatro condições diferentes. Em um primeiro momento, foi processado um cenário simples de teste, através do qual foi possível validar o funcionamento dos protótipos explicar com maiores detalhes o processo de aprendizagem.

Em um segundo momento, os protótipos processaram um cenário complexo, de forma que se pudesse verificar a convergência do algoritmo Q-learning utilizado nos protótipos. No terceiro momento, os protótipos processaram 20 cenários aéreos escolhidos aleatoriamente. O processamento destes cenários teve como finalidade o treinamento dos protótipos, ou seja, a conformação da Q-table. Finalmente os protótipos processaram um cenário "desconhecido" em condições reais de aplicação, momento em que foi possível analisar o desempenho a e aplicabilidade dos protótipos.

A seguir são apresentadas informações sobre as configurações e parâmetros de cenários utilizados no estudo de caso. Serão discutidos com detalhes os cenários de teste e aplicação. Os cenários de treinamento não serão discutidos, posto que são em número elevado e as suas características básicas são muito semelhantes aos cenários (teste e aplicação) discutidos detalhadamente.

6.2.1 Setorização do Espaço Aéreo

Conforme foi discutido previamente, o espaço aéreo brasileiro é dividido em 46 setores de controle, distribuídos em 5 Regiões de Informação de Vôo - FIR (FIR Brasília, com 12 setores; FIR Curitiba, com 10 setores; FIR Recife, com 8 setores; FIR Amazônica, com 14 setores; FIR Atlântico, com 2 setores). Contudo, de acordo com a demanda de tráfego, e segundo regras operacionais específicas, a configuração dos setores de controle pode ser alterada, mediante o agrupamento de determinados setores. Observando as regras constantes dos modelos operacionais dos órgãos de controle, o experimento adotou a seguinte configuração:

- FIR Brasília (FIR BS) - setores 1 e 2 agrupados, setores 3 e 4 agrupados, setores 7 a 9 agrupados;
- FIR Curitiba (FIR CW) - setores 1 e 2 agrupados, setores 4 e 5 agrupados, setores 6 a 8 agrupados, setores 13 e 14 agrupados;
- FIR Recife (FIR RE) - setores 1 e 2 agrupados, setores 3 e 4 agrupados;

- FIR Amazônica (FIR AZ) - setores 3 a 5 agrupados, setores 6 e 7 agrupados, setores 8 a 10 agrupados, setores 11 a 14 agrupados.

Neste ponto, cabe ressaltar que o agrupamento de setores reduz drasticamente a capacidade do espaço aéreo; por exemplo, cada setor das FIR pode comportar isoladamente 14 aeronaves simultaneamente, ao passo que agrupamentos de setores (não importa quantos setores estão agrupados) comportam 12 aeronaves simultaneamente. Neste contexto, a adoção da configuração acima teve dois objetivos principais: primeiramente, reduzir propositadamente a capacidade do espaço, forçando condições de saturação que proporcionaram um maior número de ciclos de aprendizagem; em adição, o agrupamento utilizado contribuiu para a simplificação do conjunto de estados, que se reduziu para 27^2 .

Por fim, as estatísticas de movimento nas FIR disponibilizado pelo CGNA sinalizam que o movimento mensal na FIR Atlântico (FIR AO) representa apenas 1,6% do movimento observado no espaço aéreo jurisdicionado ao Brasil, não apresentando histórico de saturação. Como consequência, os setores da FIR AO não foram considerados no experimento. Desta feita, os testes contaram com um conjunto de estados da ordem de 25^2 .

6.2.1.1 Referências do Cenário (Notação)

Os setores são referenciados da seguinte forma:

- FIR BS: setores 1 e 2 agrupados = 1.1-2; setores 3 e 4 agrupados = 1.3-4; setor 5 = 1.5; setor 6 = 1.6; setores 7, 8 e 9 agrupados = 1.7-9; setor 10 = 1.10; setor 11 = 1.11; setor 12 = 1.12;
- FIR CW: setores 1 e 2 agrupados = 2.1-2; setor 3 = 2.3; setores 4 e 5 agrupados = 2.4-5; setores 6, 7 e 8 agrupados = 2.6-8; setores 13 e 14 agrupados = 2.13-14;
- FIR RE: setores 1 e 2 agrupados = 3.1-2; setores 3 e 4 agrupados = 3.3-4; setor 5 = 3.5; setor 6 = 3.6; setor 7 = 3.7; setor 8 = 3.8;
- FIR AZ: setor 1 = 4.1; setor 2 = 4.2; setores 3, 4 e 5 agrupados = 4.3-5; setores 6 e 7 agrupados = 4.6-7; setores 8, 9 e 10 agrupados = 4.8-10; setores 11, 12, 13 e 14 agrupados = 4.11-14.

6.2.2 Fator FPL

As aeronaves que evoluem no espaço aéreo tem suas decolagens autorizadas após a apresentação e aprovação do seu plano de vôo pelos órgãos que gerenciam a utilização do espaço aéreo. As aeronaves com demandas regulares (empresas aéreas) decolam segundo planos de vôo denominados RPL (Plano de

Vôo Repetitivo), os quais são analisados e aprovados pelo Centro de Gerenciamento da Navegação Aérea, com uma antecedência de 15 dias em relação a data e horário da decolagem. Existe total controle sobre esses planos, que podem ser modificados/ajustados pelo CGNA em função da relação capacidade x demanda.

Por sua vez, as aeronaves com demandas eventuais (aviação em geral, aeronaves militares) decolam segundo planos de vôo denominados FPL (Plano de Vôo), os quais são apresentados 45 minutos antes do horário previsto para a decolagem. Tais planos ficam sujeitos a aprovação do órgão de controle de tráfego aéreo com jurisdição sobre o local de partida da aeronave. Estes planos são disponibilizados aos órgãos de controle, para análise da condição demanda x capacidade, 20 minutos antes do horário de decolagem.

O controle sobre tais planos é portanto muito limitado, não havendo um efetivo controle ou arquivamento de demandas passadas. Neste contexto, o CGNA iniciou um trabalho de análise estatística da ocorrência de FPL em áreas de maior movimento, objetivando aprimorar o planejamento (projeções, simulações) da utilização do espaço aéreo. A partir deste trabalho desenvolvido pelo CGNA, foi possível estabelecer valores percentuais médios inerentes à participação dos FPL na demanda de tráfego global. Os testes realizados neste trabalho incorporaram tais valores, doravante referenciados como Fator RPL. Considerando tais valores, as demandas definidas pelos RPL são incrementadas de valores percentuais que representam as demandas definidas pelos FPL. O valores definidos para as FIR foram os seguintes: FIR BS - 20%, FIR CW - 20%, FIR RE - 10%, FIR AZ - 5%.

6.2.3 Cenário de Teste

As Tabelas 6.2.1 e 6.2.2 referem-se ao cenário aéreo (distribuição das aeronaves nos setores de controle) do dia 06 de março de 2008 entre 08:15Hs e 13:30Hs.

Tabela 6.2.1: Previsão de Demanda RPL com Fator FPL - FIR BS / FIR CW (06/03/2008)

Hora	FIR BS							FIR CW					
	1.1-2	1.3-4	1.5	1.6	1.7-9	1.10	1.11	1.12	2.1-2	2.3	2.4-5	2.6-8	2.9-10
08:15	0	1	1	2	2	5	1	8	1	0	1	4	6
09:00	1	6	6	2	5	0	1	8	2	1	1	1	3
09:45	6	1	2	4	2	4	8	1	6	8	7	2	10
10:30	4	5	6	6	5	0	10	1	9	10	3	0	4
11:15	13	7	2	0	13	4	11	7	8	9	3	1	2
12:00	10	4	6	4	8	5	7	2	6	10	6	7	9
12:05	13	1	7	4	5	7	5	1	8	6	3	8	14
12:45	11	7	12	1	6	10	5	4	4	1	9	9	11
13:30	12	4	7	5	11	1	10	5	6	10	9	7	2

Na Tabela 6.2.1, por exemplo, às 11:55Hs os setores 1 e 2 da FIR Brasília agrupados (1.1-2) apresentam uma demanda (projetada) de 13 movimentos aéreos. Como os setores estão agrupados, o limite máximo de 12 aeronaves no setor tende a ser ultrapassado, caso nenhuma medida restritiva de fluxo de tráfego seja aplicada. Às 12:05Hs os setores 9 e 10 da FIR Curitiba agrupados (2.9-10) apresentam uma demanda (projetada) de 14 movimentos aéreos. Mais uma vez, como os setores estão agrupados, o limite máximo de 12 aeronaves no setor tende a ser ultrapassado, caso nenhuma medida restritiva de fluxo de tráfego seja aplicada.

Tabela 6.2.2: Previsão de Demanda RPL com Fator FPL - FIR RE / FIR AZ (06/03/2008)

Hora	FIR RE						FIR AZ					
	3.1-2	3.3-4	3.5	3.6	3.7	3.8	4.1	4.2	4.3-5	4.6-7	4.8-10	4.11-14
08:15	0	2	1	8	3	1	1	1	1	4	2	1
09:00	2	2	3	2	0	0	1	0	2	2	1	0
09:45	3	7	0	2	1	2	1	1	2	0	0	1
10:30	7	2	6	1	6	5	7	1	1	0	1	2
11:15	2	1	3	3	10	0	0	0	1	1	2	0
12:00	2	7	0	2	1	2	2	0	2	0	0	1
12:05	1	6	0	2	1	2	2	0	3	0	0	1
12:45	1	1	3	5	7	2	0	2	0	2	2	0
13:30	2	1	7	1	8	2	0	0	1	1	0	1

Conforme pode ser observado na Tabela 6.2.2, não há saturação nas FIR RE e FIR AZ. A partir da Tabela 6.2.1, foram projetados os seguintes pontos de saturação:

- FIR BS setores 1 e 2 agrupados (1.1-2) às 11:15Hs, contando 13 movimentos aéreos;
- FIR BS setores 7, 8 e 9 agrupados às 11:15Hs, contando 13 movimentos aéreos;

- FIR BS setores 1 e 2 agrupados às 12:05Hs, contando 13 movimentos aéreos;
- FIR BS setores 1 e 2 agrupados às 13:30Hs, contando com 12 movimentos aéreos;
- FIR CW setores 9 e 10 agrupados às 12:05Hs, contando 14 movimentos aéreos.

6.2.4 Cenário de Verificação de Convergência

As Tabelas 6.2.3 e 6.2.4 referem-se ao cenário aéreo (projeção inicial da distribuição das aeronaves nos setores de controle) do dia 18 de março de 2009 entre 10:15Hs e 12:30Hs. Este cenário é significativamente mais complexo do que o cenário anterior, apresentando um número maior de saturações, bem como saturações mais agudas.

Tabela 6.2.3: Previsão de Demanda RPL com fator FPL - FIR BS e FIR CW (18/03/2009)

Hora	FIR BS								FIR CW				
	1.1-2	1.3-4	1.5	1.6	1.7-9	1.10	1.11	1.12	2.1-2	2.3	2.4-5	2.6-8	2.9-10
10:15	13	10	6	5	11	8	4	2	20	13	9	2	2
11:00	5	8	8	4	16	11	11	5	20	8	7	2	6
11:45	19	8	6	0	11	11	4	7	10	20	9	7	12
12:30	10	11	8	2	7	5	5	6	10	3	12	6	4

Tabela 6.2.4: Previsão de Demanda RPL com fator FPL - FIR RE e FIR AZ (18/03/2009)

Hora	FIR RE						FIR AZ					
	3.1-2	3.3-4	3.5	3.6	3.7	3.8	4.1	4.2	4.3-5	4.6-7	4.8-10	4.11-14
10:15	8	8	7	4	3	7	6	1	8	4	3	1
11:00	6	4	4	2	4	7	3	1	3	0	2	2
11:45	2	4	4	3	2	3	1	1	4	1	2	2
12:30	1	7	6	2	3	0	1	1	3	1	1	3

Conforme pode ser observado na Tabela 6.2.4, não foram projetados pontos de saturação nas FIR RE e FIR AZ. A partir da Tabela 6.2.3, foram projetados os seguintes pontos de saturação:

- FIR BS às 10:15Hs: setores 1 e 2 agrupados (1.1-2), contando 13 movimentos aéreos;
- FIR BS às 11:00Hs: setores 7, 8 e 9 agrupados (1.7-9), contando 16 movimentos aéreos;

- FIR BS às 11:45Hs: setores 1 e 2 agrupados (1.1-2), contando 19 movimentos aéreos;
- FIR CW às 10:15Hs: setores 1 e 2 agrupados (2.1-2), contando 20 movimentos aéreos;
- FIR CW às 11:00Hs: setores 1 e 2 agrupados (2.1-2), contando 20 movimentos aéreos;
- FIR CW às 11:45Hs: setor 3 (2.3), setores 9 e 10 agrupados (2.9-10), contando com 20 e 12 movimentos aéreos respectivamente;
- FIR CW às 12:30Hs: setores 4 e 5 agrupados (2.4-5), contando com 12 movimentos aéreos.

6.2.5 Cenário de Aplicação

As Tabelas 6.2.5 e 6.2.6 referem-se ao cenário aéreo (projeção da distribuição das aeronaves nos setores de controle) do dia 17 de março de 2010, entre 10:15Hs e 12:30Hs.

Tabela 6.2.5: Previsão de Demanda RPL com fator FPL - FIR BS e FIR CW (17/03/2010)

Hora	FIR BS								FIR CW				
	1.1-2	1.3-4	1.5	1.6	1.7-9	1.10	1.11	1.12	2.1-2	2.3	2.4-5	2.6-8	2.9-10
10:15	10	12	2	7	22	12	10	4	19	11	12	14	9
11:00	12	11	2	11	20	12	10	6	14	12	15	12	4
11:45	18	13	10	5	12	10	10	7	10	14	15	21	12
12:30	17	16	16	6	13	6	10	2	8	11	21	12	13

Tabela 6.2.6: Previsão de Demanda RPL com fator FPL - FIR RE e FIR AZ (17/03/2010)

Hora	FIR RE						FIR AZ					
	3.1-2	3.3-4	3.5	3.6	3.7	3.8	4.1	4.2	4.3-5	4.6-7	4.8-10	4.11-14
10:15	7	8	3	8	8	9	7	4	5	7	5	5
11:00	10	4	6	6	9	4	12	2	12	4	1	6
11:45	9	6	3	7	9	4	7	4	12	12	8	5
12:30	8	4	6	6	4	2	3	3	15	12	4	14

Conforme pode ser observado nas Tabelas 6.2.5 e 6.2.6, foram projetados os pontos de saturação listados a seguir.

- FIR BS às 10:15Hs: setores 3 e 4 agrupados (1.3-4), setores 7, 8 e 9 agrupados (1.7-9), com 12 e 22 movimentos aéreos respectivamente;
- FIR BS às 11:00Hs: setores 1 e 2 agrupados (1.1-2), setores 7, 8 e 9 agrupados (1.7-9), com 12 e 20 movimentos aéreos respectivamente;

- FIR BS às 11:45Hs: setores 1 e 2 agrupados (1.1-2), setores 3 e 4 agrupados (1.3-4); setores 7, 8 e 9 agrupados (1.7-9), com 18, 13 e 12 movimentos aéreos respectivamente;
- FIR BS às 12:30Hs: setores 1 e 2 agrupados (1.1-2), setores 3 e 4 agrupados (1.3-4); setor 5 (1.5), setores 7, 8 e 9 agrupados (1.7-9), com 17, 16, 16 e 13 movimentos aéreos respectivamente;
- FIR CW às 10:15Hs: setores 1 e 2 agrupados (2.1-2), setores 4 e 5 agrupados (2.4-5), setores 6, 7 e 8 agrupados (2.6-8), com 19, 12 e 14 movimentos aéreos respectivamente;
- FIR CW às 11:00Hs: setores 1 e 2 agrupados (2.1-2), setores 4 e 5 agrupados (2.4-5), setores 6, 7 e 8 agrupados (2.6-8), com 14, 15 e 12 movimentos aéreos respectivamente;
- FIR CW às 11:45Hs: setor 3 (2.3), setores 4 e 5 agrupados (2.4-5), setores 6, 7 e 8 agrupados (2.6-8), setores 9 e 10 agrupados (2.9-10), com 14, 15, 21 e 12 movimentos aéreos respectivamente;
- FIR CW às 12:30Hs: setores 4 e 5 agrupados (2.4-5), setores 6, 7 e 8 agrupados (2.6-8), setores 9 e 10 agrupados (2.9-10), com 21, 12 e 13 movimentos aéreos respectivamente;
- FIR AZ às 11:00Hs: setores 3, 4 e 5 agrupados (4.3-5), com 12 movimentos aéreos;
- FIR AZ às 11:45Hs: setores 3,4 e 5 agrupados (4.3-5), setores 6 e 7 agrupados (4.6-7), ambos com 12 movimentos aéreos;
- FIR AZ às 12:30Hs: setores 3,4 e 5 agrupados (4.3-5), setores 6 e 7 agrupados (4.6-7), setores 11, 12, 13 e 14 agrupados (4.11-14), com 15, 12 e 14 movimentos aéreos respectivamente.

6.2.6 Parâmetros Iniciais de Aprendizagem

O algoritmo Q-learning foi configurado com os parâmetros de aprendizagem apresentados na Tabela 6.2.7.

Tabela 6.2.7: Parâmetros de Aprendizagem

Parâmetro	Valor
Reforço	$T(e) = -(\varphi D(e) + \beta C(e)), \varphi + \beta = 1$
Constantes de Normalização	$\varphi = 0,0136985, \beta = 0,9863015$
Desconto temporal	$\gamma = 0,8$
Taxa de Aprendizagem	$\alpha = 0,2$
Critério <i>Exploration / Exploitation</i>	1 a 25.000 iterações = 20% / 80%
	25.001 a 40.000 iterações = 5% / 95%
	Acima de 40.001 iterações = 1% / 99%

Sobre a Tabela 6.2.7, considera-se oportuno realizar os seguintes comentários:

- As equações que possibilitam o cálculo do Reforço foram apresentadas e discutidas na Subseção 5.3.2.
- As Constantes de Normalização φ e β foram dimensionadas de forma que cada fator, no pior caso possível, contribua com aproximadamente 50% do valor do Reforço, conforme discutido na Subseção 5.3.2.
- Os valores de γ e α foram definidos tomando por base estudo realizado (Ribeiro *et al*, 2006) para avaliar a eficiência do algoritmo Q-learning em função do número de estados, do desconto temporal (γ) e da taxa de aprendizagem (α).
- Os critérios *Exploration / Exploitation* foram definidos experimentalmente, considerados os comentários apresentados na Subseção 4.4.1.

6.3 Teste Inicial de Validação dos Protótipos

A seguir são apresentados os resultados obtidos no teste inicial de validação dos protótipos MAAD e MAAD*. Nesta fase, os protótipos processaram um cenário aéreo bastante simples (6 de março de 2008), com poucas saturações, e através do qual ilustra-se o processo de aprendizagem.

A partir do cenário apresentado na Subseção 6.2.3, foram sugeridas as ações apresentadas nas Tabelas 6.3.1 e 6.3.2. Na coluna com o *label* "MAAD" são apresentadas as ações (atrasos) sugeridas, nos horários determinados, pelo protótipo MAAD; na coluna com *label* "MAAD*" são apresentadas as ações (atrasos) sugeridas pelo protótipo MAAD*.

Tabela 6.3.1: Medidas de Controle de Fluxo - 1 experimento

HORA	10:55		11:00		11:05		11:10	
	POLÍTICA		POLÍTICA		POLÍTICA		POLÍTICA	
TMA	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*
SP	5	5	-	5RE	-	5RE	-	-
BR	5	5	5	5SP	5	5SP	5	5SP
BH	5	5	5	5SP	5	5SP	-	5SP

Conforme pode ser observado na Tabela 6.3.1, num primeiro momento (às 10:55Hs) foram geradas restrições de fluxo (atrasos em solo) apenas para as TMA-SP, TMA-BR e TMA-BH (5 minutos de atraso em todas as decolagens partindo das Terminais São Paulo (SP), Brasília (BR) e Belo Horizonte (BH)). Observa-se também que o Protótipo MAAD* gerou restrições de fluxo diferenciadas (empregando ações obtidas junto ao CGNA) já no segundo ciclo de aplicação das referidas medidas (às 11:00Hs): 5 minutos para as decolagens partindo da TMA SP para a FIR Recife (RE); 5 minutos para as decolagens partindo da TMA BR para a TMA SP; e 5 minutos para as decolagens partindo da TMA BH para a TMA SP. Neste ponto cabe lembrar que o sistema opera segundo ciclos de projeção de cenário de 5 minutos.

Tabela 6.3.2: Medidas de Controle de Fluxo - 1 experimento (continuação)

HORA	11:45		11:50		11:55		12:00	
	POLÍTICA		POLÍTICA		POLÍTICA		POLÍTICA	
TMA	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*
SP	5	5	5	-	-	7RE	-	5RE
BR	5	5	5	7	5	7SP	5	-
BH	5	5	5	7	5	5	5	-
RJ	5	5	5	5	-	-	-	-
VT	5	5	5	5	5	-	-	-
FL	-	-	5	5	5	-	-	-
SV	-	-	-	-	5	5SP	-	-

Na Tabela 6.3.2, verifica-se que os protótipos, na maior parte dos ciclos de projeção, apresentaram ações diferentes. Nos ciclos iniciados às 11:55Hs e 12:00Hs, o protótipo MAAD* sugeriu a aplicação de medidas restritivas obtidas a partir dos relatórios do CGNA. Esta condição foi compreendida como um comportamento normal do sistema, na medida em que tais ações implicam impactos pontuais e reduzidos no cenário, e considerando ainda a simplicidade do cenário.

No início do treinamento, havia uma expectativa de que os protótipos somente apresentassem saídas diferentes nos momentos em que fossem sugeridas, pelo MAAD*, ações oriundas dos relatórios do CGNA. Contudo, às 11:50Hs houve discrepância entre as medidas sugeridas pelos protótipos, e somente

foram sugeridas medidas geradas internamente pelo sistema (MBF). Em função desse resultado, foi feita uma análise da Q-table dos protótipos e verificou-se que, dado o caráter aleatório da opção *exploration / exploitation* e considerando as características do algoritmo Q-learning, as ações sugeridas podem ser diferentes, quando se considera a realização de poucos experimentos.

Em função da situação descrita previamente, e com o objetivo de se validar os protótipos, foi realizado um número mínimo de experimentos, a partir do qual se pudesse observar comportamentos semelhantes dos protótipos. Por experimento, entende-se o processamento de um mesmo cenário. Desta feita, o experimento do dia 06 de março de 2008 foi repetido 100 vezes e, ao final, foram obtidos os resultados apresentados nas Tabelas 6.3.3 e 6.3.4.

Tabela 6.3.3: Medidas de Controle de Fluxo - 100 experimentos

HORA	10:55		11:00		11:05		11:10	
	POLÍTICA		POLÍTICA		POLÍTICA		POLÍTICA	
TMA	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*
SP	5	5	-	5RE	-	5RE	-	-
BR	5	5	5	5SP	5	5SP	5	5SP
BH	5	5	5	5SP	5	5SP	-	5SP

Tabela 6.3.4: Medidas de Controle de Fluxo - 100 experimentos (continuação)

HORA	11:45		11:50		11:55		12:00	
	POLÍTICA		POLÍTICA		POLÍTICA		POLÍTICA	
TMA	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*	MAAD	MAAD*
SP	5	5	5	5	-	7RE	-	5RE
BR	5	5	5	5	5	7SP	5	-
BH	5	5	5	5	5	5	5	-
RJ	5	5	5	5	-	-	-	-
VT	5	5	5	5	5	-	-	-
FL	-	-	5	5	5	-	-	-
SV	-	-	-	-	5	5SP	-	-

Nas Tabelas 6.3.3 e 6.3.4, os resultados apresentados em cada horário representam as ações sugeridas com maior frequência, considerando os 100 experimentos. Ao final dos experimentos, o comportamento dos protótipos foi julgado adequado para a realização do teste de desempenho apresentado posteriormente.

No processo de treinamento, a cada iteração, o algoritmo Q-learning sugere uma ação (medida de ajuste de fluxo) e atualiza a Q-table. A seguir são descritos os passos observados no ciclo de aprendizagem.

1. Projeção de Cenário - através do MAPC, e com base nos RPL (incluindo o Fator FPL), são projetados os cenários onde o sistema atuará; tais projeções abarcam as decolagens previstas para os próximos 120 minutos e a quantidade de aeronaves nos setores de controle no momento, tendo como referência $t = t_{instante\ da\ previsão} + 20\ minutos$).
2. Detecção da Saturação - através do MAPC, são detectados os pontos de saturação no cenário projetado, decorrendo a ativação do MBF.
3. Composição do Conjunto de Ações - o MBF, mediante a utilização do critério de fluxo máximo (Teoria de Grafos), gera ações (esperas em solo) preliminares e as envia para o MAAD; tais ações são incorporadas ao conjunto de ações do MAAD.
4. Ajuste de Fluxo - o MAAD consulta a sua tabela de experiências (Q-table) e busca a melhor ação para o cenário em vigor ou busca, no conjunto de ações, uma ação ainda não utilizada/avaliada; a opção entre estas duas possibilidades (*exploration / exploitation*) se dá conforme especificado em 6.2.6.
5. Atualização de Cenário - o MAAD envia as ações sugeridas ao MAPC, o qual efetua a projeção do novo cenário.
6. Atualização da Base de Experiências - o MAAD atualiza a sua base de experiências (Q-table) considerando a ação aplicada e o novo cenário produzido.

Após a última fase, novas previsões de saturação são efetuadas e o ciclo se repete, cabendo ressaltar que as projeções de cenário abarcam intervalos mínimos de 5 minutos (ciclos de projeção).

6.4 Experimentos e Resultados - Agente ATFM sem a Experiência Humana

6.4.1 Verificação de Convergência - Protótipo MAAD

Nesta fase do estudo de caso, procede-se à verificação de convergência do algoritmo Q-learning empregado no protótipo MAAD, caracterizado por um conjunto de ações que não incorpora a experiência do especialista humano (gerente de fluxo do CGNA).

Com o objetivo de analisar o comportamento dos protótipos no processamento de cenários complexos (grande número de saturações agudas) e complementar as informações sobre a forma como os protótipos foram treinados,

o cenário de verificação de convergência foi processado com a utilização de uma Q-table vazia. A partir do cenário apresentado na Subseção 6.2.4, foram sugeridas as ações apresentadas na Tabela 6.4.1.

Tabela 6.4.1: MAAD - Medidas de Controle de Fluxo / Número de Iterações

HORA	100	5000	25000	50000
TMA	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)
SP	7	7	10	7
BR	10	7	7	10
BH	15	15	15	7
AN	5	0	0	0
UL	5	0	0	0
RJ	7	7	10	10
CT	10	7	7	5
PA	10	10	10	5
FL	10	10	10	5
SV	5	5	5	5
RE	5	5	5	0
FZ	5	5	5	5
MN	5	5	5	0
BE	5	5	5	0

Diferentemente do que foi observado na fase anterior, no presente cenário as ações são apresentadas em função do número de iterações, pois nesta fase, os protótipos realizaram iterações (ciclos de aprendizagem) até que a condição de parada fosse atingida, ou seja, até que o cenário não mais apresentasse saturações. Desta forma, o sistema "treina" com um número de cenários projetados igual ao número de ciclos de aprendizagem realizados. Na Tabela 6.4.1, por exemplo, ao atingir 5000 iterações, o sistema teve como saída a aplicação, dentre outros, de um atraso de 7 minutos para as aeronaves partindo das TMA SP, BR, RJ e CT.

A seguir são apresentados os valores de $Q(s,a)$ referente ao protótipo MAAD, em função do número de iterações realizadas. Portanto, a Figura 6.4.1 apresenta o comportamento do protótipo MAAD, considerando a condição de parada.

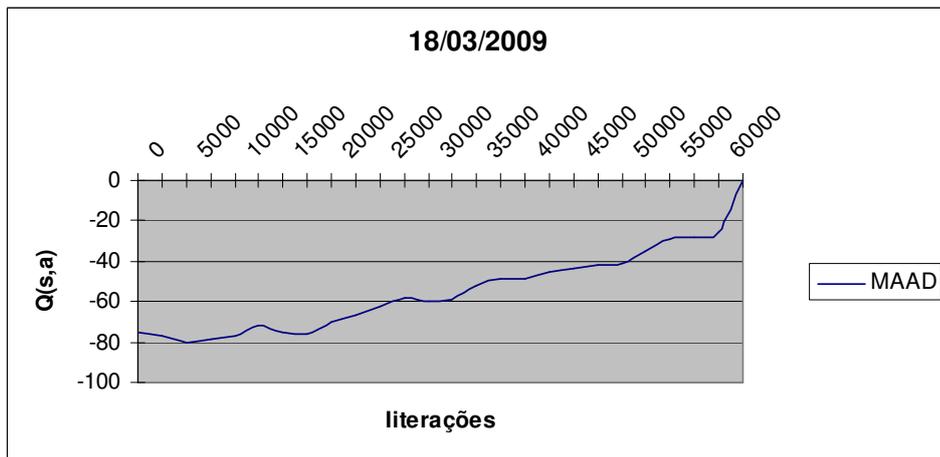


Figura 6.4.1: Convergência - Protótipo MAAD

Conforme pode ser observado na Figura 6.4.1, o protótipo apresentou um comportamento que sinaliza a convergência do algoritmo Q-learning. O MAAD levou cerca de 60.000 iterações (ciclos de aprendizagem) para alcançar a condição de parada, ou seja, os valores de $Q(s, a)$ evoluíram de -75 para aproximadamente 0, resultando em um cenário sem congestionamento.

6.4.2 Aplicação do Protótipo MAAD

Nesta fase do estudo de caso, o protótipo MAAD foi aplicado nas mesmas condições de uma utilização real no CGNA. Ou seja, após treinado, o protótipo processou um cenário "desconhecido" e sugeriu ações para ajustar o fluxo de tráfego e reduzir a quantidade de aeronaves nos setores de controle.

A partir do cenário apresentado na Subseção 6.2.5, foram sugeridas as ações apresentadas nas Tabelas 6.4.2 e 6.4.3. As medidas restritivas (atrasos) são apresentadas em intervalos de 25 minutos projetados (5 ciclos de projeção de cenário / aprendizagem).

Tabela 6.4.2: MAAD - Medidas de Controle de Fluxo - Aplicação

HORA	09:55	10:20	10:45	11:10
TMA	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)
SP	15	15	10	7
BR	15	15	15	10
BH	15	10	10	10
AN	7	-	-	-
UL	5	-	-	-
RJ	7	5	5	7
CT	10	7	5	7
PA	7	7	5	-
FL	5	5	5	-
CG	5	-	-	-
LO	5	5	-	-
SV	7	5	-	-
RE	7	5	-	-
FZ	7	-	-	-
MN	7	-	-	-
BE	7	5	-	-
ST	5	-	-	-

Na Tabela 6.4.2, por exemplo, às 10:20Hs, o sistema teve como saída a aplicação, dentre outros, de um atraso de 7 minutos para as aeronaves partindo das TMA CT (Curitiba) e PA (Porto Alegre).

Tabela 6.4.3: MAAD - Medidas de Controle de Fluxo - Aplicação (continuação)

HORA	11:35	12:00	12:25
TMA	ATRASO (MINUTO)	ATRASO (MINUTO)	ATRASO (MINUTO)
SP	7	5	5
BR	10	7	7
BH	7	7	7
AN	-	-	-
UL	-	-	-
RJ	5	-	-
CT	5	7	-
PA	5	-	-
FL	5	-	-
LO	-	-	-
CG	-	-	-
SV	-	-	-
RE	-	-	-
FZ	-	-	-
MN	-	-	-
BE	-	-	-
ST	-	-	-

O gráfico da Figura 6.4.2 apresenta o comportamento dos valores de $Q(s,a)$

inerentes ao protótipo MAAD. Maiores comentários sobre tal comportamento constam da Seção 6.6.

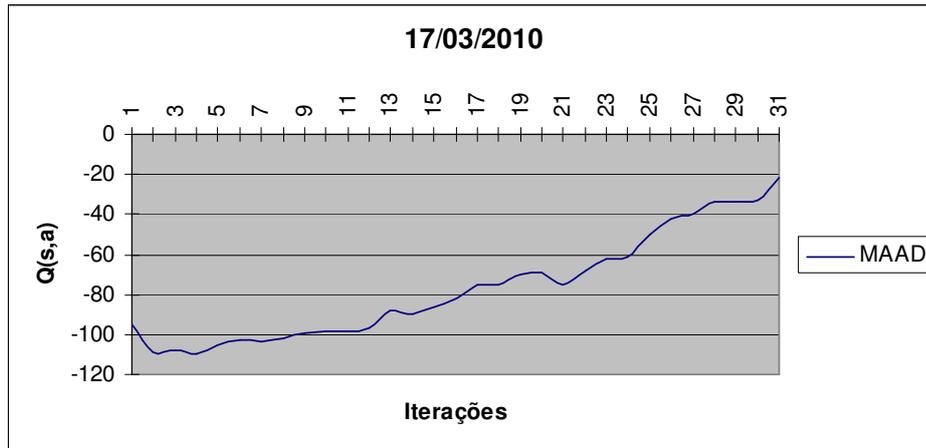


Figura 6.4.2: Valores de $Q(s,a)$ - MAAD

6.5 Experimentos e Resultados - Agente ATFM com a Experiência Humana

6.5.1 Verificação de Convergência - Protótipo MAAD*

Nesta fase do estudo de caso, procede-se à verificação de convergência do algoritmo Q-learning empregado no protótipo MAAD*, caracterizado por um conjunto de ações que incorpora a experiência do especialista humano (gerente de fluxo do CGNA). Considerando as informações de cenário contidas na Subseção 6.2.4, referentes à projeção inicial da distribuição das aeronaves nos setores de controle, no dia 18 de março de 2009, o MAAD* sugeriu as ações apresentadas na Tabela 6.5.1.

Tabela 6.5.1: MAAD* - Medidas de Controle de Fluxo / Número de Iterações

HORA	100	5000	25000	50000
TMA	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)	POLÍTICA MAAD*
SP	10	7	7	7-RE
BR	7	7	7	7-SP
BH	15	15	10	7-SP
AN	5	-	-	-
UL	5	-	-	-
RJ	10	10	7	-
CT	7	7	7	5
PA	10	10	10	5
FL	10	7	10	5
SV	7	7	5	-
RE	5	5	5	-
FZ	5	5	5	-
MN	5	5	5	-
BE	5	5	5	-

Na Tabela 6.5.1, por exemplo, ao atingir 25000 iterações, o sistema teve como saída a aplicação, dentre outros, de um atraso de 10 minutos para as aeronaves partindo das TMA BH (Belo Horizonte), PA (Porto Alegre) e FL (Florianópolis).

A Figura 6.5.1 apresenta o comportamento do protótipo MAAD*, considerando a condição de parada. Conforme pode ser observado, o protótipo apresentou um comportamento que sinaliza a convergência do algoritmo Q-learning. O MAAD* levou mais de 70.000 iterações (ciclos de aprendizagem) para alcançar a condição de parada, ou seja, os valores de $Q(s, a)$ evoluíram de -79 para 0.

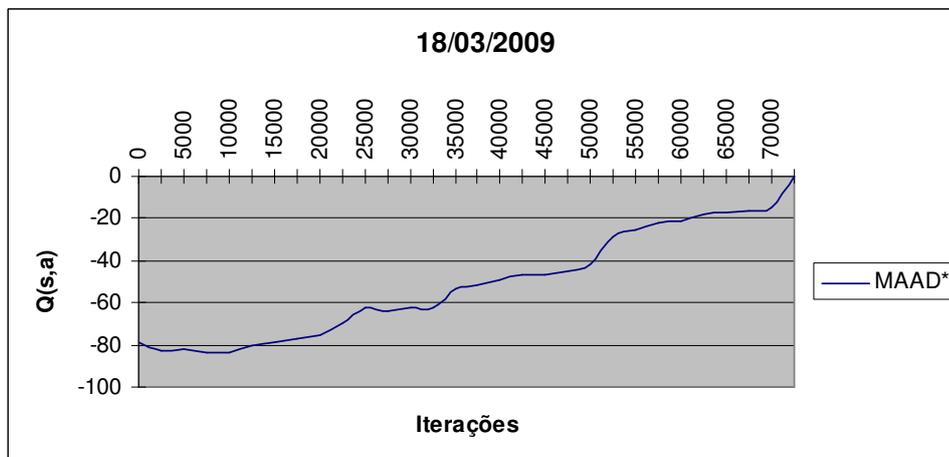


Figura 6.5.1: Convergência - Protótipo MAAD*

6.5.2 Aplicação do Protótipo MAAD*

A partir do cenário apresentado na Subseção 6.2.5, o protótipo MAAD* sugeriu as ações apresentadas nas Tabelas 6.5.2 e 6.5.3. As medidas restritivas são apresentadas em intervalos de 25 minutos projetados (5 ciclos de projeção de cenário / aprendizagem). O comportamento dos protótipos na aplicação pode ser observado no gráfico da Figura 6.5.2.

Tabela 6.5.2: MAAD* - Medidas de Controle de Fluxo - Aplicação

HORA	09:55	10:20	10:45	11:10
TMA	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)	ATRASO (MINUTOS)
SP	15	10	10	7
BR	15	15	10	10
BH	15	15	10	10
AN	5	5	-	-
UL	5	5	-	-
RJ	10	5	5	5
CT	10	7	5	5
PA	7	7	5	-
FL	7	5	5	-
CG	5	-	-	-
LO	5	-	-	-
SV	7	7	-	-
RE	5	5	-	-
FZ	5	-	-	-
MN	5	-	-	-
BE	7	5	-	-
ST	5	-	-	-

Tabela 6.5.3: MAAD* - Medidas de Controle de Fluxo - Aplicação

HORA	11:35	12:00	12:25
TMA	ATRASO (MINUTO)	ATRASO (MINUTO)	ATRASO (MINUTO)
SP	7	7BR	5BR
BR	10	5SP	5SP
BH	7	5SP	5SP
AN	-	-	-
UL	-	-	-
RJ	5	-	-
CT	5	5SP	-
PA	5	5SP	-
FL	5	5SP	-
LO	-	-	-
CG	-	-	-
SV	-	-	-
RE	-	-	-
FZ	-	-	-
MN	-	-	-
BE	-	-	-
ST	-	-	-

Na Tabela 6.5.3, por exemplo, às 12:25Hs, o sistema teve como saída a aplicação, dentre outros, de um atraso de 5 minutos para as aeronaves partindo das TMA BR (Brasília) e BH (Belo Horizonte) com destino a TMA SP. O gráfico da Figura 6.5.2 apresenta o comportamento dos valores de $Q(s,a)$ inerentes ao protótipo MAAD. Maiores comentários sobre tal comportamento constam da Seção 6.6.

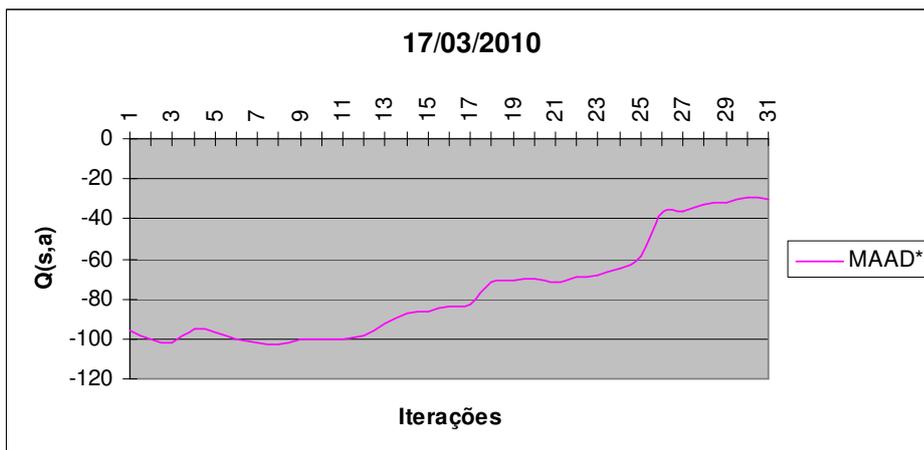


Figura 6.5.2: Valores de $Q(s,a)$ - MAAD*

6.6 Discussão de Resultados

6.6.1 Tempo de Execução

Com o objetivo de avaliar o tempo de execução dos protótipos, desenvolvidos em linguagem Java (Java versão 1.5 / banco de dados relacional MySQL 5.0.45), os testes foram conduzidos mediante a utilização de duas configurações distintas de *hardware* e *software*:

- Estação de trabalho 1 dotada com: a) processador Intel Core2Duo T5250 (*clock* 1.5Ghz); b) memória RAM de 2Gb; c) sistema operacional Windows Vista 32 bits.
- Estação de Trabalho 2 dotada com: a) processador Intel Core2Quad Q8300 (*clock* 2,5Ghz; memória RAM de 4Gb; c) sistema operacional Fenix Extreme 2.0.

Os tempos de execução observados constam da Tabela 6.6.1. Os tempos são apresentados no formato HH:MM:SS.

Tabela 6.6.1: Tempo de Aprendizagem

N. de iterações	Estação 1		Estação 2	
	MAAD	MAAD*	MAAD	MAAD*
1	00:00:12	00:00:14	00:00:12	00:00:13
1000	01:45:35	01:59:43	01:42:05	01:44:35
5000	09:17:20	12:03:55	08:53:45	10:54:25
50000	45:43:00	51:20:20	42:12:03	43:05:24

Os resultados observados na Tabela 6.6.1 sinalizam que os tempos de execução verificados não impõem restrição à aplicação dos protótipos, consideradas as condições de realização do estudo de caso, com destaque para as reduções impostas ao conjunto de estados, tal como discutido nas Subseções 5.2.4 e 6.2.1.

6.6.2 Análise da Convergência

A Figura 6.6.1 apresenta o comportamento dos protótipos observado na fase de verificação de convergência, em um único gráfico, de forma a facilitar uma análise comparativa.

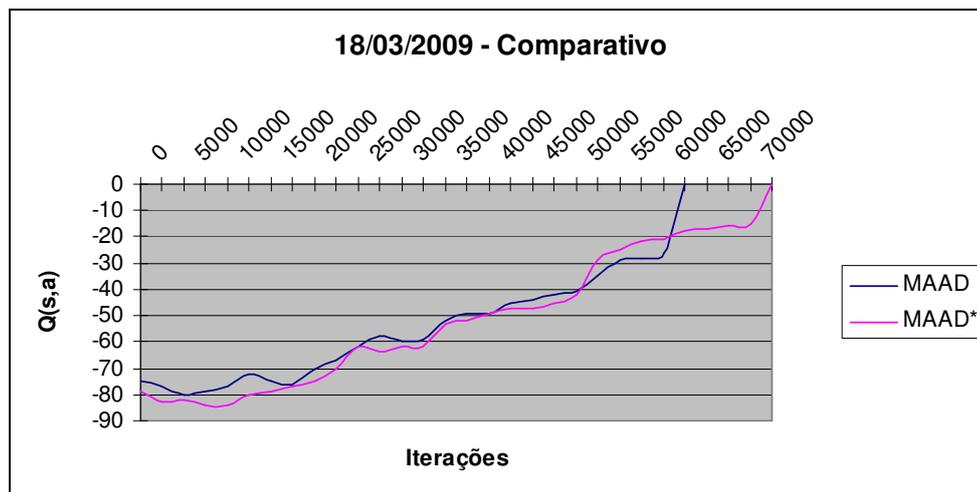


Figura 6.6.1: Convergência - Gráfico Comparativo

Sobre a verificação da convergência e o treinamento dos protótipos, julga-se oportuno tecer os seguintes comentários:

1. Os dois principais objetivos da fase de treinamento foram: a) montar o conjunto de ações do protótipo MAAD e parte do conjunto de ações do protótipo MAAD*; b) montar uma Q-table que possibilitasse a redução do número de iterações necessárias à convergência do Q-learning, de forma a otimizar o desempenho dos protótipos na fase de aplicação; e c) verificar o comportamento (convergência) do algoritmo Q-learning.
2. Os ciclos de projeção de cenário (5 minutos) dizem respeito à projeção da evolução dos movimentos aéreos nos cenários, em função das medidas sugeridas, considerando um ambiente de simulação em tempo acelerado. Desta feita, o tempo de execução dos protótipos se define basicamente pelo número de iterações realizadas (ciclos de aprendizagem).
3. Conforme pode ser observado na Figura 6.6.1, o protótipo MAAD assume tendência de convergência a partir de aproximadamente 5000 iterações.
4. Conforme pode ser observado na Figura 6.6.1, o protótipo MAAD* assume tendência de convergência a partir de aproximadamente 10000 iterações.
5. Conforme pode ser observado na Figura 6.6.1, a partir de 50000 iterações, o protótipo MAAD* passa a apresentar um desempenho ligeiramente superior ao apresentado pelo MAAD. Contudo, o MAAD* requer um número maior de iterações para atingir a condição de parada.
6. A melhora de desempenho do MAAD*, a partir de 50.000 iterações, pode ser atribuída a utilização das ações produzidas pelos gerentes de fluxo

(mais pontuais); ou seja, conforme expectativas iniciais, os gerentes de fluxo conseguem conceber ações eficientes (menor impacto negativo no cenário) ao analisar cenários menos complexos.

7. A maior quantidade de iterações demandadas pelo MAAD*, para alcançar o estado-meta, pode ser explicada por duas condições: pela maior dimensão da Q-table; e pelo caráter menos impactante das ações produzidas pelos gerentes de fluxo, implicando um maior período de aplicação das medidas de controle. Não foi possível isolar as contribuições desses fatores.
8. Na fase de treinamento, o sistema processou 20 cenários (cada qual com duração entre 2 e 5 horas), tais como os apresentados até o momento, e gerou 150 ações (conjuntos de medidas restritivas ou atrasos em solo), as quais compuseram o conjunto de ações do protótipo MAAD. A este conjunto de ações foram acrescidas 44 ações, selecionadas, segundo o critério de recorrência, a partir dos Relatórios de Situação Diária do CGNA, compondo-se assim o conjunto de ações do protótipo MAAD*.

Uma vez concluído o treinamento e conformada a Q-table, foi possível comparar o comportamento dos protótipos em uma aplicação semelhante a aplicação real; a análise comparativa pode ser melhor conduzida a partir da observação do gráfico apresentado na Figura 6.6.2.

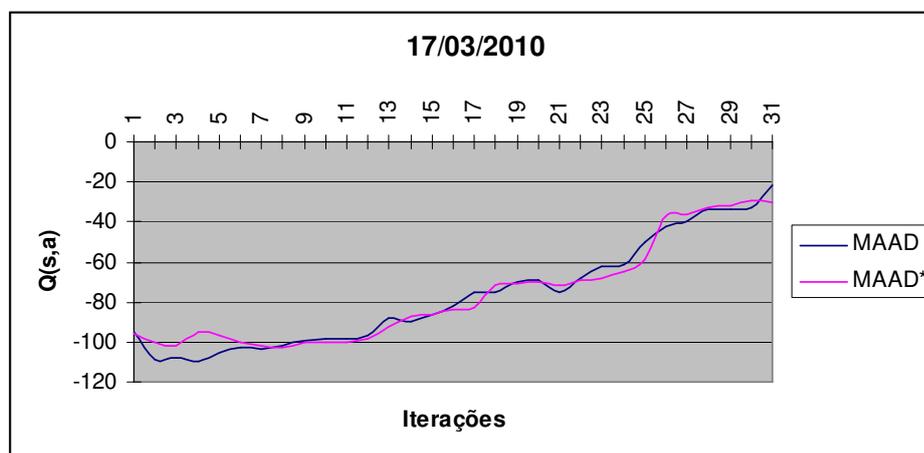


Figura 6.6.2: Comportamento - MAAD x MAAD*

Sobre a aplicação dos protótipos, julga-se oportuno tecer os seguintes comentários:

1. Utilizando a Q-table e os conjuntos de ações gerados na fase de treinamento, os protótipos processaram o cenário em tela observando as condições de aplicação real, ou seja, a cada iteração, o algoritmo Q-learning sugeriu

uma ação e um novo cenário foi projetado considerando tal ação; tal ciclo foi repetido pelo número de vezes correspondente ao número de ciclos de projeção (5 minutos) contidos no intervalo de tempo de duração do cenário.

- Os valores apresentados na Figura 6.6.2 representam os valores médios de $Q(s,a)$, para os dois protótipos, ao final de 100 repetições do experimento (processamento completo do cenário). Cada experimento contou com 31 iterações, ressaltando que a cada iteração uma ação é sugerida e um novo cenário é projetado. Conforme pode ser observado, os protótipos apresentaram comportamento muito semelhantes. Contudo, o protótipo MAAD* sugeriu medidas mais pontuais, caracterizadas por impactar apenas aeronaves com destino a uma determinada FIR/TMA.

6.6.3 Alteração dos Parâmetros de Aprendizagem

Com o objetivo de ampliar a análise do comportamento dos protótipos, foi realizada uma verificação de convergência considerando novos parâmetros de aprendizagem. Para a realização desta análise ampliada, foram adotados os parâmetros de aprendizagem apresentados na tabela a seguir.

Tabela 6.6.2: Parâmetros de Aprendizagem

Parâmetro	Valor
Reforço	$T(e) = -(\varphi D(e) + \beta C(e)), \varphi + \beta = 1$
Constantes de Normalização	$\varphi = 0,0136985, \beta = 0,9863015$
Desconto temporal	$\gamma = 0,6$
Taxa de Aprendizagem	$\alpha = 0,1$
Critério <i>Exploration / Exploitation</i>	1 a 25.000 iterações = 20% / 80%
	25.001 a 40.000 iterações = 5% / 95%
	Acima de 40.001 iterações = 1% / 99%

Com a utilização dos novos parâmetros de aprendizagem, os valores de $Q(s,a)$, inerentes ao protótipo MAAD, apresentaram o comportamento observado no gráfico da Figura 6.6.3.

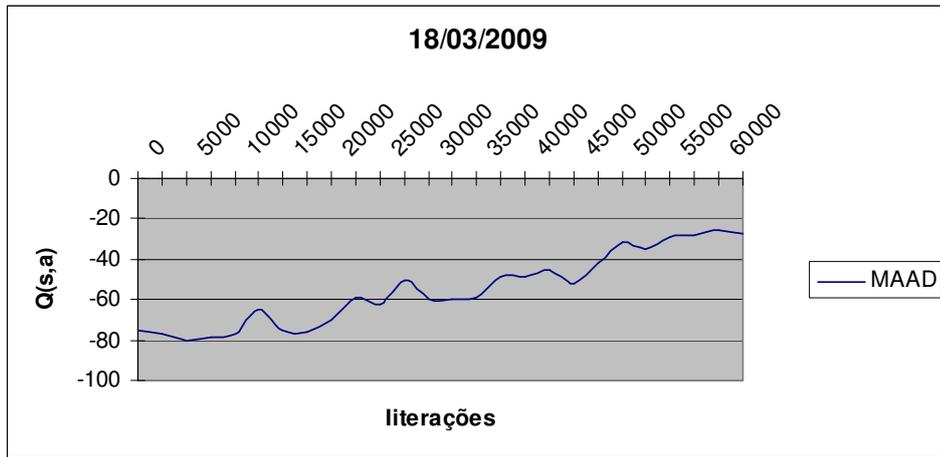


Figura 6.6.3: MAAD - $Q(s,a)$

Ao aplicar os novos parâmetros no protótipo MAAD*, observa-se o seguinte comportamento dos valores de $Q(s,a)$.

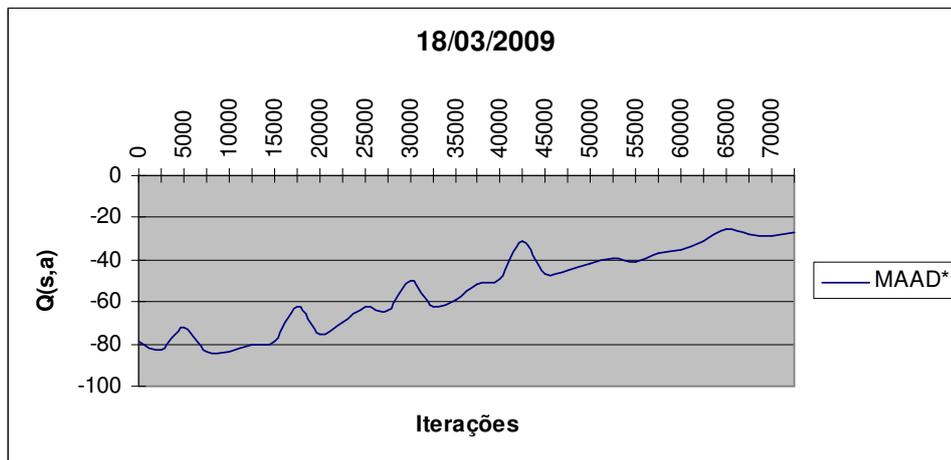


Figura 6.6.4: MAAD* - $Q(s,a)$

Ao analisar o comportamento dos valores de $Q(s,a)$ em ambos os protótipos, torna-se possível tecer os seguinte comentários:

- O processo de aprendizagem torna-se mais lento, posto que o número de ciclos necessários para o alcance da condição de parada, para ambos os protótipos, foi superior ao observado com os parâmetros de aprendizagem iniciais; tal condição se explica principalmente pela redução do valor da taxa de aprendizagem α .
- Verifica-se que os protótipos tendem a buscar ações com alto valor imediato de recompensa, o que se explica pelos vários picos nos valores observados; essa condição, contudo, não contribui para o melhor desempenho dos protótipos no longo prazo; tal condição pode ser explicada pela redução no fator de desconto γ .

6.6.4 Comentários Finais

A apresentação dos resultados obtidos no estudo de caso abarcou três partes principais, através das quais foram gerados resultados para três cenários distintos. A primeira parte, considerando o processamento de um cenário simples (poucas saturações), permitiu verificar o correto funcionamento do protótipo. Nesta parte foram elencados os passos inerentes ao ciclo de aprendizagem dos protótipos.

Na segunda parte ocorreu o processamento de um cenário de treinamento mais complexo, caracterizado por elevado número de saturações. Nesta parte do estudo de caso, verificou-se o comportamento dos protótipos na fase de treinamento propriamente dita, através do acompanhamento dos valores de $Q(s, a)$ em função do número de iterações realizadas até que fosse atingida a condição de parada.

Pela análise dos valores de $Q(s, a)$, foi possível verificar que os protótipos apresentaram comportamentos sensivelmente diferentes, e que tais diferenças tiveram duas causas principais, a saber: a) diferença na quantidade de ações nos respectivos conjuntos de ações, com a conseqüente diferença no tamanho das Q-tables; e b) característica das ações incorporadas a partir dos relatórios do CGNA, constatando-se que tais ações são mais pontuais, causando menos impacto (menores valores de $D(e)$); contudo, tais ações tendem a perdurar por um maior período de tempo (menor grau de convergência).

A fase de treinamento completa consistiu do processamento de 20 cenários semelhantes ao cenário do dia 18 de março de 2010. Ao final do treinamento, os protótipos MAAD e MAAD* geraram conjuntos de ações com 150 e 194 ações diferentes respectivamente, assim como as respectivas Q-tables (tabelas de experiência do agente). Passou-se então ao processamento do cenário de aplicação e, para tanto, foi escolhido o cenário do dia 17 de março de 2010.

Com o processamento do cenário referente à demanda do dia 17 de março de 2010, foi possível verificar a capacidade dos protótipos de sugerir medidas com ação efetiva sobre cenários com elevada quantidade de saturações. Os crescentes valores de $Q(s, a)$, mostrados na Figura 6.6.2, corroboram tal afirmação. Em adição, as medidas de controle de fluxo de tráfego apresentadas na Tabela 6.5.3 indicam que a agregação do MAAD* ao SISCONFLUX representa uma importante contribuição na busca da solução para uma questão crucial no contexto da problemática ATFM, qual seja, a necessidade de se aplicar medidas de controle de fluxo mais pontuais, afetando apenas aeronaves com determinados

destinos.

A aplicação dos protótipos foi conduzida da forma como se daria a aplicação real dos mesmos. Assim sendo, os valores de $Q(s, a)$ foram apresentados (e analisados) a cada iteração realizada, considerando a projeção de um novo cenário a partir da ação sugerida pelo Q-learning. Nesta fase, os protótipos apresentaram ligeira diferença de comportamento, conforme pode ser observado no gráfico comparativo, mas mantendo padrões satisfatórios de convergência. Os resultados obtidos nesta fase sinalizam que os protótipos desenvolvidos podem ser de grande utilidade ao apoiar o processo decisório ATFM. Por fim, verificou-se que os parâmetros α e γ inicialmente sugeridos foram os mais adequados para a utilização dos protótipos.

Capítulo 7

CONCLUSÃO

No âmbito do Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB), a atividade de Gerenciamento de Fluxo de Tráfego Aéreo (ATFM) se reveste de fundamental importância, mormente quando considerados dois aspectos de extrema relevância, quais sejam: primeiramente, o impacto do ATFM (tático) na atividade de controle de tráfego aéreo, incluindo as inerentes implicações na segurança operacional; e, num segundo momento, as possíveis consequências das medidas ATFM sobre a logística aeroportuária. Assim sendo, torna-se imperativo o estabelecimento de processos e/ou o desenvolvimento de sistemas (agentes computacionais) que auxiliem os gerentes de fluxo de tráfego (agentes humanos) a adotar ações otimizadas.

Neste contexto, o presente trabalho apresentou a materialização da proposta de desenvolvimento de um agente computacional (inteligente) baseado em aprendizagem por reforço que tem por objetivo sugerir atrasos nos horários de decolagem das aeronaves partindo de determinada TMA, de forma que os setores de controle de tráfego aéreo não saturem, em função de um desbalanceamento eventual e momentâneo entre demanda e capacidade. O trabalho incluiu um estudo de caso onde foram comparadas as ações sugeridas pelo agente de forma autônoma e as medidas sugeridas levando em consideração a experiência dos gerentes de fluxo de tráfego aéreo lotados no Centro de Gerenciamento da Navegação Aérea (CGNA).

Os resultados obtidos no estudo de caso são promissores, na medida em que o comportamento dos agentes concebidos indicam que o Algoritmo Q-learning convergiu de forma satisfatória, decorrendo que os protótipos geraram ações que contribuíram de forma efetiva para a redução das saturações nos cenários aéreos testados. Desta feita, com o desenvolvimento dos protótipos MAAD e MAAD*, foi possível demonstrar a aplicabilidade de formalismos da Inteligência Artificial na busca de soluções nacionais, para a Problemática ATFM, inteira-

mente concebidas segundo as características e necessidades do país.

Contudo, algumas restrições importantes foram estabelecidas com o intuito de viabilizar a utilização dos protótipos, em função da complexidade de espaço e de tempo inerente ao sistema desenvolvido. Por exemplo, foi necessário reduzir a dimensão do conjunto de estados, mediante a aplicação de dois artifícios: a) representação da quantidade de aeronaves nos setores de controle de forma simplificada (saturado / não saturado); b) processamento de cenários com uma reduzida quantidade de setores de controle, mediante o agrupamento de setores segundo diretrizes e regras em vigor nos centros de controle de tráfego aéreo. Assim sendo, a efetiva utilização do sistema concebido requer o prosseguimento das pesquisas, de forma que se busque soluções que permitam processar o cenário aéreo de forma integral e observando complexidade de tempo e estado adequados ao tempo de resposta requerido pelas demandas ATFM.

7.1 Contribuições e Pesquisas Futuras

Após a conclusão do presente trabalho, é possível delinear de forma objetiva as contribuições agregadas ao estudo realizado, quais sejam:

- a) Através do desenvolvimento dos protótipos do MAAD, foi possível contribuir para a implementação de um sistema de gerenciamento de fluxo de tráfego aéreo voltado para o incremento da eficiência das medidas de controle de fluxo de tráfego aplicadas no país;
- b) O MAAD agregou um processo de aprendizagem por reforço com capacidade de incorporar a experiência dos especialistas humanos, quais sejam, os gerentes de fluxo de tráfego aéreo do CGNA;
- c) O MAAD agregou ao SISCONFLUX a capacidade de gerar conjuntos de medidas de controle de fluxo de tráfego segundo critérios pré-estabelecidos, a saber: montante de atraso imputado aos movimentos aéreos afetados pelas medidas e nível de congestionamento dos setores de controle de tráfego aéreo.

Para a efetivação das contribuições supracitadas, foram desenvolvidas as seguintes atividades

- a) Definição da arquitetura do Módulo de Avaliação e Apoio à Decisão (MAAD) do SISCONFLUX;
- b) Concepção e aplicação de uma função de avaliação de cenário (estado) que permitiu calcular as recompensas utilizadas pelo algoritmo Q-learning, considerando os critérios (congestionamento de

setores de controle e montante de atraso produzido pela medida de controle) estabelecidos;

- c) Desenvolvimento e aplicação de um agente inteligente, baseado em aprendizagem por reforço, concebido para sugerir medidas de controle de fluxo de tráfego aéreo e com capacidade para incorporar a experiência dos referidos gerentes;
- d) Estabelecimento de um estudo comparativo acerca do desempenho do agente considerando duas condições distintas: atuação do agente independentemente da experiência dos gerentes de fluxo e atuação do agente considerando a incorporação da experiência dos gerentes de fluxo.

A partir das contribuições elencadas, sugere-se o desenvolvimento de pesquisas futuras:

- a) O desenvolvimento de agentes inteligentes com capacidade para sugerir outras medidas de controle de fluxo de tráfego aéreo, tais como espera em voo e rotas alternativas;
- b) O desenvolvimento de sistemas de gerenciamento de fluxo de tráfego empregando técnicas multi-agentes, com o objetivo de reduzir a dimensão do conjunto de ações, do conjunto de estados e da Q-table, de forma a garantir escalabilidade ao sistema;
- c) O desenvolvimento de interfaces entre os diversos sistemas de apoio à decisão empregados no ATFM, objetivando a padronização e o compartilhamento das bases de dados.

7.2 Considerações Finais

Com a finalização do MAAD, consolida-se a proposta de implementação do SIS-CONFLUX, com aplicabilidade dirigida ao ATFM tático. Tal projeto constitui objeto de cooperação entre o Comando da Aeronáutica e a Universidade de Brasília, com vistas ao desenvolvimento de um sistema de extrema importância no contexto operacional dos órgãos de controle de tráfego e que elidirá uma importante lacuna dentro da concepção nacional de gerenciamento do tráfego aéreo, devendo ser empregado em conjunto com as ferramentas computacionais de apoio a decisão existentes (STVD, SYNCROMAX, et.). Desta feita, o SIS-CEAB poderá dispor de um conjunto de ferramentas capaz de suportar o gerenciamento de fluxo de tráfego nos seus três níveis: estratégico, pré-tático e tático.

A equipe de projeto apresenta um caráter multidisciplinar e é integrada por pesquisadores da UnB, do CINDACTA I e da Atech, contando com o apoio do Centro de Gerenciamento da Navegação Aérea, órgão do Departamento de Controle do Espaço Aéreo responsável pelo planejamento e execução do gerenciamento de tráfego aéreo no Brasil.

Referências Bibliográficas

- [1] ALVES, D. P. (2006). Modelagem usando Aprendizagem por Reforço em Sistemas Multi-agentes para um Ambiente Controlado em Nível Meta. Dissertação de mestrado, Universidade de Brasília - UnB, Brasília, DF, Brasil.
- [2] ALVES, D. P., WEIGANG, L., SOUZA, B. B. (2006). Using Meta-Level Control with Reinforcement Learning to Improve the Performance of the Agents. In: International Conference on Fuzzy Systems and Knowledge Discovery - FSKD, 2006, Xi An. LNAI - International Conference on Fuzzy Systems and Knowledge Discovery - FSKD. Berlin Heidelberg: Springer-Verlag, 2006. v. 4223. p. 1109-1113.
- [3] ALVES, D. P., WEIGANG, L., SOUZA, B. B. (2008). Reinforcement Learning to Support Meta-Level Control in Air Traffic Management. In: Cornelius Weber, Mark Elshaw and Norbert Michael Mayer. (Org.). Reinforcement Learning - Theory and Applications. Vienna: ARS publishing, 2008, v. 1, p. 409-424.
- [4] AGOGINO, A., TUMER, K. (2008). Regulating Air Traffic Flow with Coupled Agents. Proceedings of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008). Padgham, Parkes, Müller and Parsons (eds.). Estoril, Portugal. pp. 535-542.
- [5] AGOGINO, A., TUMER, k. (2009). Learning Indirect Actions in Complex Domains: Action Sugestions for Air Traffic Control. Advances in Complex Systems, Volume: 12, Issues: 4-5, 2009, 493-512 World Scientific Company.
- [6] ARRUDA, A. C. (2009). Aprendizagem por Reforço Aplicada à Análise de Impacto no Controle de Fluxo de Tráfego Aéreo. Dissertação de mestrado, Universidade de Brasília - UnB, Brasília, DF, Brasil.
- [7] BALL, M. O., HOFFMAN, R., ODoni, A., AND RIFKIN, R. (2003). Stochastic Integer Program with Dual Network Structure and its Application to the Ground Holding Problem. Operations Research Volume 51 , Issue 1, pp. 167-171. Institute for Operations Research and the Management Sciences (INFORMS), Linthicum, Maryland, USA. ISSN:0030-364X

- [8] BAYEN, A. M., GRIEDER, P., MEYER, G., e TOMLIN, C. J. (2005). Lagrangian Delay Predictive Model for Sector-based Air Traffic Flow, *AIAA Journal of Guidance, Control, and Dynamics* 28 (2005) pp. 1015-1026.
- [9] BERTSIMAS, D., PATTERSON, S. S. (1998). The Air Traffic Management Problem with Enroute Capacities. *Operations Research* 46, pp. 406-422.
- [10] BERTISIMAS, D., LULLI, G., ODONI, A. (2008). The Air Traffic Management Problem: An Integer Optimization Approach. *IPCO 2008, LNCS 5035*, pp. 34-46. Springer-Verlag Berlin Heidelberg.
- [11] CRESPO, A. M. F., AQUINO, C. V., SOUZA, B. B., WEIGANG, L., MELO, A. C. M. A., ALVES, D. P. (2007). Sistema Distribuído de Apoio à Decisão Aplicado Ao Gerenciamento Tático do Fluxo de Tráfego: Caso CINDACTA I, em *Anais do VI Simpósio de Transporte Aéreo - SITRAER*, pp. 317-327, Maringá.
- [12] DECEA (2007). ICA 100-22 Serviço de Gerenciamento de Fluxo de Tráfego Aéreo. (www.aisweb.aer.mil.br). Acesso em Julho/2009.
- [13] DELL'OLMO, P. e LULLI, G. (2003). A Dynamic Programming Approach for the Airport Capacity Allocation Problem, *IMA Journal of Management Mathematics* 14 (2003), 235–249.
- [14] DIB, M. V. P. (2005). Sistema para Sincronização e Gerenciamento de Fluxo de Tráfego Aéreo Multi-agentes. *Revista Transportes*, vol. XIII, n.1, pp. 48-56. ANPET.
- [15] EVANS, L. C. e GARIEPY, R. (1992). *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics, CRC Press.
- [16] FAA (2001). The Measure of Air Traffic Control Sector Complexity for the En Route Environment: Phase II Experiment Plan. (http://www.tc.faa.gov/acb300/330_documents_dd.asp). Acessado em julho/2009.
- [17] ICAO (2005). DOC 9854: Global Air Traffic Management Operational Concept. DOC 9854-AN/458. International Civil Aviation Organization.
- [18] INFRAERO (2009). Movimentos nos Aeroportos. <http://www.infraero.gov.br>. Acesso em julho/2009
- [19] KAELBLING, L. P., LITTMAN, M. L., MOORE, A. W. (1996). Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research* 4:237-285.

- [20] LIN, L. (1992), Self-Improving Reactive Agents Based On Reinforcement Learning, Planning and Teaching. Machine Learning. Boston, MA, USA: Kluwer Academic Publishers, 1992, vol. 8, pp. 69-97.
- [21] MA, Z., CUI, D., E CHENG, P. (2004). Dynamic Network Flow Model for Short-term Air Traffic Flow Management. IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, 34(3):351–358. Work supported by National Natural Science Foundation of China.
- [22] MIYAMARU, D. O. (2008) Tendências na Evolução dos Sistemas de Gerenciamento de Tráfego Aéreo no Brasil In: VII Simpósio de Transporte Aéreo, 2008, Rio de Janeiro. Anais do VII Simpósio de Transporte Aéreo, p. XII-XVIII.
- [23] MUKHERJEE, A. (2004), Dynamic Stochastic Optimization Models for Air Traffic Flow Management, Ph.D. thesis, University of California, Berkeley, 2004, UCB-ITS-DS-2004-3.
- [24] NASA (1996). An Evaluation of Air Traffic Control Complexity. Ames Research Center. Final Report. Contract Number NAS2-14284.
- [25] RIBEIRO, R., KOERICH, A. L., ENEMBRECK, F. (2006). Uma Nova Metodologia para Avaliação de Desempenho de Algoritmos Baseados em Aprendizagem por Reforço. In: XXXIII Seminário Integrado de Software e Hardware, 2006, Campo Grande-MS. Anais do XXVI Congresso da Sociedade Brasileira de Computação.
- [26] RIZZI, J. A. (2003). Um modelo matemático de auxílio para o problema de controle do tráfego aéreo. Master's thesis, Instituto Tecnológico de Aeronáutica - ITA, São José dos Campos, SP, Brasil.
- [27] RUSSEL, S. e NORVIG, P. (2003). Artificial Intelligence – A modern Approach. Pearson Education, Inc., Second Edition, New Jersey.
- [28] SOUZA, B.B., WEIGANG, L., CRESPO, A. M. F. (2008). Implementação de modelo para balanceamento de fluxos aplicado ao gerenciamento de tráfego aéreo. Revista Transportes - (ANPET), v. 16, p. 48-58.
- [29] SRIDHAR, B., SHETH, K. S.; GRABBE, S. (1998). Airspace Complexity and its Application in Air Traffic Management. 2nd USA/Europe Air Traffic Management R&D Seminar. Orlando.
- [30] STOLTZ, S. e KY, P. (2001). Reducing Traffic Bunching More Flexible Air Traffic Flow Management, 4th USA/Europe ATM R&D Seminar, New Mexico.

- [31] STOLTZ, S., GUERREAU, R. (2002). EUROCONTROL: Future ATFM Measures (FAM) operational Concept, EEC Note No. 13/02.
- [32] SUTTON, R. S. (1988). Learning to Predict by Methods of Temporal Differences. *Machine Learning*, 3:9-44.
- [33] SUTTON, R. S., BARTO, A. G. (1998). *Reinforcement Learning: An Introduction*. The MIT Press, England.
- [34] TIMOSZCZUK, A. P., PIZZO, W. N., STANISCIA, G. F., SIEWERDT, E. (2009). The SYNCROMAX Solution for Air Traffic Flow Management in Brazil. In: Li Weigang, Alexandre de Barros, Italo Romani de Oliveira. (Org.). *Computational Models, Software Engineering, and Advanced Technologies in Air Transportation: Next Generation Applications*. Hershey: IGI Global, pp. 23-37.
- [35] TUMER, K. e AGOGINO, A. (2007). Distributed agent-based air traffic flow management, in *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems (Honolulu, HI)*, pp. 330-337.
- [36] WEIGANG, L., ALVES, C. J. P., AND OMAR, N. (1997). An expert system for Air Traffic Flow Management. *Journal of Advanced Transportation*, Vol. 31, pp. 343-361. ISSN: 0197-6729.
- [37] WICKENS, C. D., HOLLANDS, J. G. (1999). *Engineering Psychology and Human Performance*. (3a. ed) Prentice-Hall.
- [38] WOLF, S. R. (2007) Supporting Air Traffic Flow Management with Agents. *American Association for Artificial Intelligence Spring Symposium: Interaction Challenges for Intelligent Assistants*.
- [39] ZHANG, Z., GAO, W. e WANG, L. (2005). Short-term flow management based on dynamic flow programming network, *Journal of the Eastern Asia Society for Transportation Studies* 6 (2005), 640–647.