UNIVERSIDADE FEDERAL DO RIO GRANDE- FURG
CURSO DE GESTÃO EM OPERAÇÕES E LOGÍSTICA

TRABALHO DE CONCLUSÃO DE CURSO

BRUNO MAX BARRETO BARROSO

A Utilização da Inteligência Artificial no Apoio à Decisão em Jogos de Guerra.

PÓS-GRADUAÇÃO *LATO SENSU*

RIO DE JANEIRO, RJ
2023

TERMO DE AUTORIZAÇAO DE USO E APROVAÇÃO

BRUNO MAX BARRETO BARROSO

A Utilização da Inteligência Artificial no Apoio à Decisão em Jogos De Guerra

Autorizo que o presente artigo científico apresentado ao Curso de Pós-Graduação *Lato Sensu* da FURG, como requisito parcial para obtenção do certificado de Especialista em Gestão de Operações e Logística, e aprovado pelos professores responsáveis pela orientação e sua aprovação, seja utilizado para pesquisas acadêmicas de outros participantes deste ou de outros cursos, a fim de aprimorar o ambiente acadêmico e a discussão entorno das temáticas aqui propostas.

A UTILIZAÇÃO DE INTELIGÊNCIA ARTIFICIAL NO APOIO À DECISÃO EM JOGOS DE GUERRA

AUTOR: Bruno Max Barreto Barroso
ORIENTADOR: Prof. Dr. André Andrade Longaray
COORIENTADOR: Prof. Paulo Roberto da Silva Munhoz

RESUMO

O estudo sobre a utilização de ferramentas de Inteligência Artificial (IA) em jogos de guerra para apoio à decisão em combate tem se tornado cada vez mais relevante com a aprofundamento e desenvolvimento de ferramentas de IA, pois a nação que detiver este conhecimento terá vantagem na velocidade em que conseguirá tomar decisões acertadas, desbalanceando assim o poder de combate em um cenário de conflito. Este trabalho tem como propósito analisar o emprego dos jogos de guerra dentro para apoio à decisão e através de uma comparação entre diferentes agentes de IA, indicar qual o caminho que essa tecnologia tomará nesse ramo. Para tal, foi realizada uma pesquisa sistemática dos trabalhos mais recentes sobre o assunto para entender como as grandes nações estão pesquisando o tema.

PALAVRAS-CHAVE: Inteligência Artificial, Jogos de Guerra e Apoio à Decisão.

A UTILIZAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO APOIO À DECISÃO EM JOGOS DE GUERRA

Autor[1], Bruno Max Barreto Barroso

RESUMO - O estudo sobre a utilização de ferramentas de Inteligência Artificial (IA) em jogos de guerra para apoio à decisão em combate tem se tornado cada vez mais relevante com a aprofundamento e desenvolvimento de ferramentas de IA, pois a nação que detiver este conhecimento terá vantagem na velocidade em que conseguirá tomar decisões acertadas, desbalanceando assim o poder de combate em um cenário de conflito. Este trabalho tem como propósito analisar o emprego dos jogos de guerra dentro para apoio à decisão e através de uma comparação entre diferentes agentes de IA, indicar qual o caminho que essa tecnologia tomará nesse ramo. Para tal, foi realizada uma pesquisa sistemática dos trabalhos mais recentes sobre o assunto para entender como as grandes nações estão pesquisando o tema.

PALAVRAS-CHAVE: Inteligência Artificial, Jogos de Guerra e Apoio à Decisão

[1] brunomb09@gmail.com

1 INTRODUÇÃO

No contexto da guerra moderna, os conflitos têm alcançado um grau elevado de complexidade, incluindo-se aos domínios terrestre, marítimo e aeroespacial, o espectro eletromagnético e o cibernético, sintetizado no conceito de combate multi-domínio (Feickert, 2021). Dessa forma, congregando diversos tipos de atitudes dentro de uma mesma campanha de forma sucessiva ou até mesmo simultânea, como ofensiva, defensiva e interagências, nos mais diversos ambientes, desde as florestas tropicais até os mais populosos centros urbanos. A quantidade de informações que os decisores militares devem processar e ponderar no momento de expedir suas ordens é considerável, o que torna o ciclo de decisão lento, gerando uma brecha que pode ser explorada pela força inimiga. O conceito de brecha pode ser entendido como os pontos de fragilidade no dispositivo inimigo, podendo fazer referência tanto a algum material ou a um posicionamento no campo de batalha (BRASIL, 2020).

Nesse ínterim surge uma corrida para o domínio da tecnologia da Inteligência Artificial (IA) em sua utilização no apoio à decisão em jogos de guerra que simulam o ambiente de um combate real.

Dada a sua capacidade de processar uma grande quantidade de dados de forma a desenvolver uma solução eficiente do ponto de vista probabilístico, a IA aplicada à jogos de guerra, tem se tornado uma opção cada vez mais viável, não só pela acurácia com que podem simular situações reais, mas também pela drástica redução de custos para treinamento do planejamento e execução de operações de grande vulto.

Diversos estudos obtiveram sucesso com a aplicação da IA em jogos de estratégia que envolvem certo nível de complexidade em suas decisões, como descrevem Vinyals et al.(2019) que criaram uma IA que aprendeu a jogar *StarCraft II* e conseguiu chegar ao *ranking* de grão-mestre competindo com jogadores humanos. E também Goecks et al.(2021b) que, no jogo *Minecraft*, conseguiram criar uma IA que permitiu combinar aprendizado por *feedback* humano com engenharia de conhecimento para resolver problemas de tarefas hierárquicas. A maior limitação que a IA enfrenta são os ambientes em que se depara com informações incompletas e com o risco inerente das ações a empreender. Kase et al. (2022) apostam em um modelo de "*Warfighter-Machine*" ou Combatente-Máquina, no qual através de uma interface, a IA possa interagir com o combatente de forma a assessorar sua decisão,

ao passo que aprende como agir em situações de incerteza analisando as ações do decisor dada as condições em que decide.

Nesse contexto, o desenvolvimento dessa capacidade tecnológica como define Schumpeter (1942) aplicada ao planejamento de operações de guerra, mais em particular no que tange ao apoio à decisão, tanto durante o planejamento, quanto no controle das ações em curso, é importante multiplicador de poder de combate e se traduz em um tema que pode desequilibrar decisivamente o poder de combate entre forças contendoras, cabendo a resolução da problemática de como estruturar um modelo de apoio à decisão com IA alinhada com a doutrina militar vigente.

Existem duas linhas de ação possíveis para resolver esse problema: adquirir o sistema de um outro país e sofrer as consequências de não dominar a tecnologia e depender sempre dessa nação para manutenção e apoio à operação do mesmo, ou desenvolver a tecnologia e se posicionar de forma igualitária com as nações mais desenvolvidas tecnologicamente. A primeira opção é mais rápida, entretanto leva o país a uma possível utilização superficial do sistema e um investimento econômico elevado, já o segundo leva mais tempo e exige a formação de pessoal qualificado para desenvolver a tecnologia, entretanto a acumulação de capacidade de inovação coloca o Brasil em posição de independência e possivelmente exportador dessa tecnologia o que alavanca o crescimento do país.

Dessa forma, o objetivo geral deste estudo é analisar como um agente de IA pode auxiliar no planejamento de uma operação e qual melhor se adequa ao Processo de Planejamento Militar (PPM). Para isso os objetivos específicos são:

1) Identificar onde melhor se aplica o apoio à decisão da IA no PPM;
2) Examinar a melhor estratégia de aprendizado para o problema; e
3) Reconhecer o agente de IA que melhor atende ao problema proposto.

Nesse contexto, o trabalho é viável, pois não exige meios extraordinários e de difícil acesso para sua elaboração. Quanto à relevância, se justifica pela grande atenção que o assunto tem recebido em todas as mídias com a disponibilização do *ChatGPT*, IA desenvolvida pela *OpenIA*, de forma gratuita para o público. Com relação a oportunidade, se justifica, pois dada a discussão sobre a regulamentação e limitação do desenvolvimento da IA, é possível que sejam impostos ao Brasil limitações ao desenvolvimento dessa tecnologia antes mesmo que tenha alcançado seu potencial máximo, ficando este reservado apenas aos países que o conquistaram primeiro.

A pesquisa tem um enfoque quantitativo com delineamento experimental, no qual é analisado o desempenho do agente de IA baseado em aprendizagem desenvolvido para apoio à decisão quanto à sua capacidade de superar os modelos reativos vigentes. Para tal, foi utilizada análise comparativa pelo método estatístico dos índices alcançados pelos agentes em suas interações dentro do jogo de guerra.

Este artigo está dividido em 5 seções. Na 1ª realizamos uma breve introdução sobre a temática abordada e os objetivos do trabalho. Na 2ª seção é descrita uma revisão de literatura e a fundamentação teórica sobre os principais conceitos discutidos, na 3º seção é explicado o delineamento metodológico do artigo. Já a 4ª seção é destinada à descrição dos resultados e na 5ª seção descreve-se as conclusões sobre o estudo.

2 REVISÃO DE LITERATURA

A presente seção de revisão de literatura tem como objetivo realizar uma revisão sistemática dos estudos relacionados à utilização da inteligência artificial no apoio à decisão em jogos de guerra. Calaça (2020) afirma que a combinação da inteligência artificial, a tomada de decisão estratégica e os jogos de guerra são um campo de pesquisa relevante e em expansão, com implicações significativas para o desenvolvimento de estratégias militares eficazes que podem levar a uma grande rapidez na tomada de decisão acelerando o Ciclo de Boyd, comprometendo o processo decisório do inimigo e consequentemente o sobrepujando.

Hammond (2001) conta que o Ciclo de Boyd foi um termo cunhado pelo Coronel *John Boyd* da Força Aérea Americana que ao analisar os combates entre os F-86 norte americanos contra os Mig-15 soviéticos na guerra da Coreia, concluiu que apesar de os Mig-15 serem aviões superiores aos F-86, duas características faziam com que os pilotos americanos pudessem ter uma consciência situacional e velocidade de reação um pouco superiores, devido ao formato arredondado do “nariz” da aeronave e os controles hidráulicos respectivamente, fazendo com que percorressem o ciclo mais rapidamente. Ele dividiu o processo do ciclo, que também ficou conhecido como OODA, em 4 passos: Observar, Orientar, Decidir e Agir. Segundo ele, quando conseguimos realizar mais rápido que o inimigo, tornamos nossas ações imprevisíveis, quebrando sua coesão mental e o forçando a iniciar um novo ciclo pela observação a fim de entender o campo de batalha.

2.1 FUNDAMENTAÇÃO TEÓRICA

2.1.1 Inteligência Artificial

De acordo com Russel e Norvig (2016), a inteligência artificial pode ser definida como "o estudo de agentes inteligentes, isto é, entidades que percebem o ambiente por meio de sensores e agem nele através de atuadores". Eles enfatizam a importância do comportamento inteligente em relação à capacidade de atingir objetivos e se adaptar a diferentes circunstâncias.

Já Nilsson (1998) argumenta que a inteligência artificial se concentra na criação de sistemas computacionais capazes de realizar tarefas que exigem inteligência humana, como compreensão da linguagem natural, reconhecimento de padrões e aprendizado de máquina. Ele destaca a necessidade de representar o conhecimento e utilizar algoritmos para solucionar problemas de forma eficiente.

Por sua vez, Kurzweil (2005) destaca o aspecto evolutivo da inteligência artificial, afirmando que seu objetivo final é desenvolver sistemas capazes de igualar ou superar a inteligência humana em todas as suas manifestações. Ele ressalta que a IA abrange áreas como redes neurais, algoritmos genéticos e processamento de linguagem natural, contribuindo para a criação de sistemas cada vez mais sofisticados.

2.1.2 Agente de Inteligência Artificial

Segundo os autores Russell e Norvig (2022), um agente de inteligência artificial é uma entidade capaz de perceber seu ambiente por meio de sensores e agir sobre ele por meio de atuadores, buscando atingir um objetivo específico.

Esse conceito de agente está fortemente ligado ao campo do Aprendizado de Máquina, onde os agentes são projetados para aprender a tomar decisões de forma autônoma a partir de experiências passadas e interações com o ambiente. Essa capacidade de aprendizado permite que o agente melhore seu desempenho ao longo do tempo, refinando suas ações com base nos resultados obtidos.

2.1.3 Aprendizado de Máquina (*Machine Learning*)

Com isso, pode-se compreender a ideia de *Machine Learning*, que é a capacidade de um sistema aprender e melhorar seu desempenho em uma determinada tarefa por meio da experiência adquirida a partir dos dados.

Bishop (2006) afirma que *machine learning* é um termo genérico que engloba uma diversidade de algoritmos e técnicas que pretendem desenvolver uma inteligência artificial para que alcance resultados a partir de um treinamento com dados conhecidos ao invés de ser explicitamente programado decisão por decisão.

Uma outra definição importante é a de Mitchell (1997, p.2): "Um programa de computador é dito aprender a partir da experiência E em relação a uma tarefa T e uma medida de desempenho P, se o seu desempenho em T, medido por P, melhorar com a experiência E."

Essa definição implica que um algoritmo de aprendizado de máquina deve ser capaz de adaptar-se e generalizar a partir dos dados que recebe, sem a necessidade de programação explícita para cada situação. Por exemplo, um algoritmo de aprendizado de máquina pode aprender a reconhecer rostos humanos em imagens, depois de ser treinado com vários exemplos de rostos.

2.1.4 Redes Neurais (Neural Network)

Segundo Silva (2023), uma rede neural é uma estrutura matemática inspirada no cérebro humano que pode ser usada para aproximar funções complexas. Uma rede neural consiste em camadas de neurônios artificiais que são conectados uns aos outros por sinapses ponderadas. As entradas são alimentadas na rede neural e passam pelas camadas ocultas até chegar à camada de saída, onde a saída final é gerada.

2.1.5 Ator- Crítico

Segundo Sutton e Barton (2018), ator crítico é um algoritmo de aprendizado por reforço que combina elementos do método do gradiente de política e do método de diferença temporal (TD). O ator crítico é composto por dois componentes principais: o ator e o crítico. O ator é responsável por escolher ações com base na política atual, enquanto o crítico avalia a política atual e fornece *feedback* para o ator.

2.1.6 Aprendizado Profundo (*Deep Learning*)

De acordo com os autores Lecun e Bengio (2015), podemos entender o *Deep Learning* como um conjunto de técnicas de *Machine Learning* baseadas em redes

neurais artificiais profundas, que são capazes de aprender e representar automaticamente características complexas e abstratas dos dados.

Afirmam ainda que essas técnicas são caracterizadas por modelos de redes neurais profundas compostos por múltiplas camadas de “neurônios”, que são conectados em uma arquitetura em cascata. Cada camada de neurônios extrai características dos dados de entrada e as passa para a próxima camada, permitindo a aprendizagem de representações hierárquicas cada vez mais complexas. Essa abordagem de múltiplas camadas é fundamental para o processamento de dados de alta dimensionalidade, como imagens, áudio e texto, em que as características relevantes podem estar distribuídas em níveis diferentes de abstração.

O *Deep Learning* tem sido amplamente aplicado com sucesso em diversas áreas, como visão computacional, processamento de linguagem natural, reconhecimento de fala, entre outros. Essas técnicas têm demonstrado capacidade para realizar tarefas complexas, superando abordagens tradicionais de *Machine Learning* em termos de desempenho e generalização.

2.1.7 Aprendizado Profundo por Reforço (Deep Reinforcement Learning)

Segundo os autores Sutton e Barto (2018), o *Deep Reinforcement Learning* refere-se ao uso de redes neurais profundas em combinação com algoritmos de aprendizado por reforço para resolver problemas complexos de tomada de decisão sequencial. Nesse contexto, o aprendizado por reforço é uma abordagem em que um agente autônomo aprende a tomar ações em um ambiente para maximizar uma recompensa cumulativa ao longo do tempo. O agente interage com o ambiente, recebe *feedback* na forma de recompensas e usa essas informações para aprender uma política ou estratégia de ação ótima.

Os autores ainda afirmam que o *Deep Reinforcement Learning* se destaca por utilizar redes neurais profundas para representar a função Q, que é usada para estimar a qualidade de uma ação em um determinado estado. As redes neurais profundas são capazes de aprender representações complexas dos estados e ações, permitindo uma generalização mais eficaz em problemas de alta dimensionalidade.

Essa abordagem tem sido aplicada com sucesso em diversos domínios, incluindo jogos, robótica, controle de sistemas, entre outros, onde é necessário tomar decisões sequenciais em ambientes complexos. O uso de redes neurais

profundas permite que o agente aprenda políticas mais sofisticadas e alcance desempenho superior em comparação com métodos de aprendizado por reforço tradicionais.

2.1.8 Função Q

Segundo Sutton e Barto (2018), a função Q é uma forma de estimar o valor de uma ação em um determinado estado, ou seja, o retorno esperado que um agente pode obter ao executar essa ação. A função Q pode ser usada para guiar o agente na escolha da melhor ação em cada estado, de acordo com uma política ótima. A política ótima é aquela que maximiza o valor esperado da recompensa total ao longo do tempo. Para encontrar a política ótima, o agente precisa aprender os valores da função Q para cada par estado-ação, o que pode ser feito por meio de algoritmos de aprendizado por reforço, como o Q-learning.

No contexto do Q-learning, a função Q pode ser representada da seguinte forma:

$$Q\left(S_t, A_t\right) = Q\left(S_t, A_t\right) + \alpha\left[R_{t+1} + \gamma\ max\ Q\left(S_{t+1}, a\right) - Q\left(S_t, A_t\right)\right]$$

Onde:

$Q\left(S_t, A_t\right)$ é o valor Q estimado para o par estado-ação no tempo t;

α é o coeficiente de aprendizado;

R_{t+1} é a recompensa recebida no tempo t+1;

γ é o fator de desconto, que representa a importância das recompensas futuras; e

$max\ Q\left(S_{t+1}, a\right)$ é o maior valor Q para o próximo estado no tempo t+1.

Ainda segundo os autores, essa regra atualiza o valor Q usando uma média ponderada entre o valor Q anterior e o novo valor Q calculado com base na recompensa e no valor máximo do próximo estado. O coeficiente de aprendizado determina o quanto o valor Q é atualizado a cada iteração do algoritmo. Um coeficiente de aprendizado alto significa que o valor Q é mais sensível às novas experiências, enquanto um coeficiente de aprendizado baixo significa que o valor Q é mais estável e depende mais das experiências passadas. O coeficiente de aprendizado pode ser constante ou variar ao longo do tempo, dependendo da situação.

2.1.9 Processo de Decisão de Markov (MDP)

Conforme Silva (2018), os processos de decisão de Markov (MDPs) são uma ferramenta matemática usada para modelar problemas de tomada de decisão sequencial em ambientes estocásticos. Um MDP é definido por um conjunto de estados, um conjunto de ações, uma função de transição que descreve a probabilidade de transição entre estados e uma função de recompensa que descreve as recompensas recebidas pelo agente em cada estado .

Silva (2018) afirma ainda que a probabilidade de o sistema estar no estado i no período (n+1) depende somente do estado em que o sistema está no período n. Ou seja, para os processos de Markov, só interessa o estado imediato. O processo de decisão de Markov é um processo de recompensa de Markov com as ações que podem ser tomadas.

Uma possível expressão matemática para representar um processo de decisão de Markov é a seguinte:

$$P\left(X_{t+1} = x | X_t = x_{t,} X_{t-1} = x_{t-1}, \dots, X_0 = x_0\right) = P\left(X_{t+1} = x | X_t = x_t\right)$$

Essa expressão significa que a probabilidade de o sistema estar no estado x no tempo t+1 depende apenas do estado xt no tempo t, e não dos estados anteriores. Essa é a propriedade de Markov, que caracteriza esse tipo de processo estocástico.

2.1.10 Algoritmo Q-Learning

De acordo com Sutton e Barto (2018), o algoritmo Q-Learning é um dos algoritmos mais simples e populares para aprendizado por reforço. Ele usa uma tabela Q para armazenar os valores Q para cada par estado-ação. O valor Q representa o retorno esperado que o agente pode obter ao tomar uma determinada ação em um determinado estado e seguir uma política ótima a partir daí. O algoritmo Q-Learning atualiza a tabela Q usando a função Q.

2.1.11 *Deep Q-Network Algorithm (DQN Algorithm)*

O algoritmo DQN foi desenvolvido por Mnih et al. (2015) para resolver uma ampla gama de jogos de Atari (alguns em nível super-humano) combinando

aprendizado por reforço e redes neurais profundas em escala. O algoritmo DQN usa uma rede neural para aproximar a função Q, que estima o valor esperado de cada ação em cada estado. A rede neural recebe como entrada o estado atual do agente, que pode ser uma imagem da tela do jogo, e produz como saída os valores Q para cada ação possível. O agente então escolhe a ação com o maior valor Q, usando uma política “epsilon-gulosa” para explorar o ambiente. O algoritmo DQN usa uma memória de repetição para armazenar as transições observadas pelo agente, e usa amostras aleatórias dessa memória para treinar a rede neural usando o método de gradiente descendente. O algoritmo DQN também usa uma rede neural alvo, que é uma cópia da rede neural principal que é atualizada periodicamente, para estabilizar o treinamento e evitar oscilações nos valores Q.

O algoritmo DQN foi aprimorado por vários trabalhos posteriores, que introduziram variantes como o Double DQN, o Dueling DQN, o Prioritized Experience Replay, e o Rainbow. Essas variantes visam resolver alguns dos problemas do DQN original, como a tendência a superestimar os valores Q, a falta de representação da estrutura dos valores Q, a falta de priorização das experiências mais importantes, e a dificuldade de combinar diferentes melhorias.

O algoritmo DQN tem aplicações em diversos domínios, como jogos, robótica, controle e navegação. Ele permite ao agente aprender a partir da sua própria interação com o ambiente, sem necessidade de supervisão ou conhecimento prévio. Ele também permite ao agente lidar com estados complexos e contínuos, como imagens ou sinais sonoros, usando redes neurais profundas para extrair as características relevantes.

2.1.12 *Prior Knowledge Deep Q-Network Algorithm (*Algoritmo PK-DQN*)*

O algoritmo PK-DQN foi proposto por Sun et al. (2020) para resolver um problema de decisão inteligente em um ambiente de jogo de guerra, usando conhecimento a priori e redes neurais profundas.

A principal diferença entre o algoritmo PK-DQN e o algoritmo DQN original é que o PK-DQN usa conhecimento a priori para melhorar o aprendizado por reforço. O conhecimento a priori é um conjunto de regras ou heurísticas que descrevem as características do problema e as estratégias ótimas ou sub ótimas para resolvê-lo. Por exemplo, no problema do jogo de guerra, o conhecimento a priori pode incluir informações sobre os tipos de unidades, as relações de força, os objetivos da

missão, as táticas de combate, etc. O algoritmo PK-DQN usa esse conhecimento a priori para inicializar os valores Q da rede neural, para guiar a exploração do agente, e para avaliar o desempenho do agente. Dessa forma, o algoritmo PK-DQN pode acelerar a velocidade de convergência e a estabilidade do aprendizado por reforço, e alcançar melhores resultados do que o algoritmo DQN.

3 Delineamento de Pesquisa

3.1 Quanto ao propósito

A pesquisa trata-se de uma revisão sistemática de bibliografia, respeitando-se os seguintes critérios de inclusão: estudos publicados nos últimos 5 anos em idioma inglês, focados na aplicação de inteligência artificial no apoio à decisão em jogos de guerra.

Excluímos estudos não relacionados ao tema, estudos duplicados e aqueles que não estavam disponíveis na íntegra. A estratégia de busca foi realizada nas bases de dados IEEE Xplore, ACM Digital Library, MDPI Open Access Journals, Cornell University Library, SAGE Journals e Frontiers Journals. As seguintes palavras-chave foram utilizadas em combinação: '*Artificial Intelligence*', '*decision support*' e '*wargaming*'. A busca foi limitada aos títulos, resumos e palavras-chave dos artigos.

A seleção dos estudos foi realizada em duas etapas. Na primeira etapa, os títulos e resumos dos artigos foram avaliados de acordo com os critérios de inclusão. Na segunda etapa, os artigos relevantes foram selecionados para uma análise detalhada. Foram extraídas as seguintes informações dos artigos selecionados: algoritmos de inteligência artificial utilizados, métodos de apoio à decisão empregados, características dos jogos de guerra utilizados, resultados obtidos e quaisquer outras informações relevantes para a pesquisa.

3.2 Quanto ao delineamento

A pesquisa tem um enfoque quantitativo com delineamento experimental, no qual é analisado o desempenho de diversos agentes de IA para apoio à decisão quanto à sua capacidade de superar os modelos reativos vigentes.

3.3 Quanto à técnica de coleta

Foram utilizados índices e relatórios escritos com os resultados obtidos por outros pesquisadores dentro de seus testes com os agentes de IA.

3.4 Quanto à técnica de análise

Foi utilizada análise comparativa pelo método estatístico dos índices alcançados pelos agentes em suas interações dentro do jogo de guerra, para estabelecer dentro das condições de um conflito real, qual IA poderia obter melhores resultados ou estaria no caminho certo para tal.

4 Resultados e Discussões

4.1 O apoio à decisão e o Processo de Planejamento Militar (PPM)

Segundo o manual EGN-181 da Escola de Guerra Naval, a utilização dos jogos de guerra tem como finalidade instruir os militares para que possam ter uma experiência mais próxima com a realidade. A utilização de jogos de guerra como apoio à decisão em operações reais, não é um assunto regulamentado no âmbito da Marinha do Brasil (MB). O que se busca em um futuro próximo é conseguir um alto nível de assessoramento para a tomada de decisão utilizando-se de jogos de guerra aliado a um agente de inteligência artificial.

Diversas outras forças armadas possuem o mesmo entendimento com relação à utilização de jogos de guerra em apoio ao desenvolvimento do planejamento de operações. Kase et al. (2022) afirmam que dentro do exército americano tem-se utilizado os jogos de guerra tanto para a confecção das *Course of Action* (COA), que seriam as Linhas de Ação (LA) dentro da doutrina da Marinha do Brasil (MB), quanto para o *wargaming,* que dentro na MB se equivale ao confronto, que consiste no embate entre as possíveis LA de nossas forças e as possibilidades do inimigo (PI), visando elencar as melhores LA para a próxima etapa do planejamento.

O Processo de Planejamento Militar (PPM) descrito no manual EMA 331 Vol.1 de 2006, tem 3 etapas que seguem a seguinte ordem: 1ª Etapa - Exame da Situação, 2ª Etapa - Desenvolvimento do Plano de Ação e Elaboração da Diretiva (DEPAED) e 3ª Etapa - Controle da Ação Planejada.

A 1ª Etapa, Exame da Situação está dividida em 6 fases, conforme pode ser visto na Figura número.

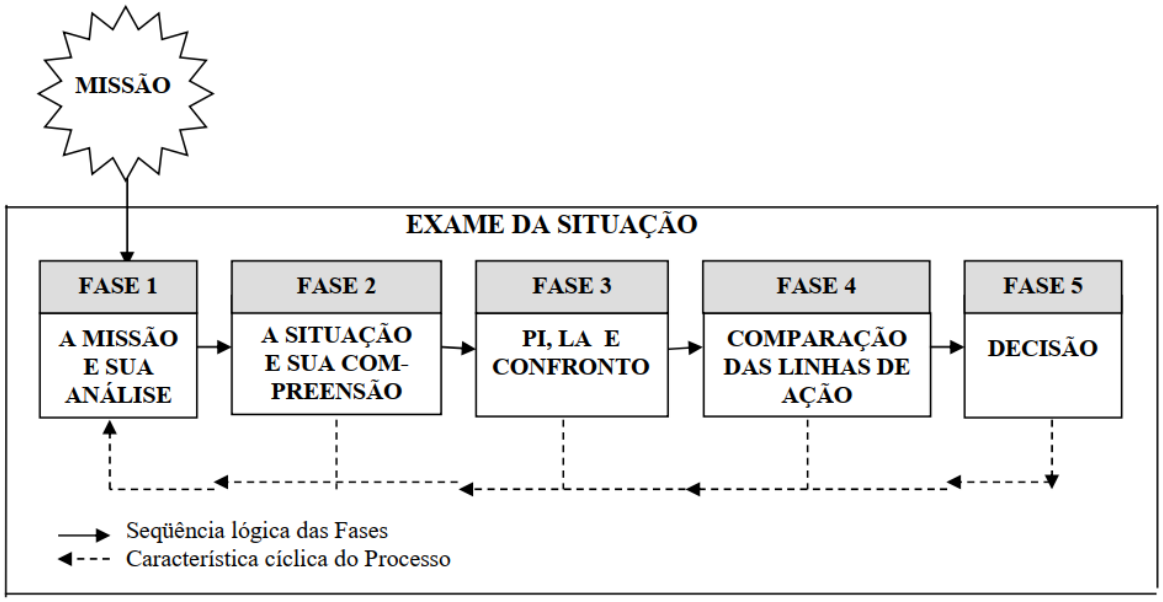

Figura 1 – Fases da 1ª Etapa do PPM
Fonte: - EMA 331 Vol. 1

No contexto da aplicação em jogos de guerra a fase 3 da 1ª etapa ganha maior relevância pois agrega os principais pontos em que a IA pode ser melhor empregada dado o volume de dados a serem processados a fim de realizar o seu treinamento, como informações do terreno, localização e capacidades do inimigo, condições climáticas, dentre outras e a partir de seu treinamento, determinar quais ações tomar de forma a alcançar o resultado esperado, provendo assim ao decisor Linhas de Ação a serem adotadas.

4.2 O ambiente do jogo de guerra e a estratégia de aprendizagem

O artigo de Zhang e Xue (2020) apresenta um método baseado em ator-crítico para a tomada de decisão do comandante de inteligência artificial em jogos de guerra táticos. Para testar o método, os autores desenvolveram um jogo de guerra como ambiente de pesquisa, chamado ArmorCombat. Segundo os autores, o jogo é baseado em um cenário de combate entre dois países fictícios, chamados

Redland e Blueland, que estão em conflito por causa de recursos naturais. O jogo contém um *script* de inteligência artificial embutido e suporta o modo de combate máquina-máquina, no qual dois agentes controlam as forças militares de cada país.

Os autores descrevem as características do jogo com o mapa dividido em células hexagonais, que representam diferentes tipos de terreno, como planície, montanha, floresta, rio, etc. Cada tipo de terreno tem um efeito sobre o movimento e o ataque das unidades militares. As unidades militares são classificadas em quatro tipos: infantaria, veículo blindado, artilharia e helicóptero. Cada tipo de unidade tem diferentes atributos, como vida, dano, alcance, velocidade, etc. As unidades podem se mover e atacar dentro do seu alcance em cada turno.

O objetivo do jogo é eliminar todas as unidades inimigas ou capturar a base inimiga. A base é uma célula especial que gera novas unidades a cada turno. Cada país tem uma base no seu território. O jogo usa um sistema de pontuação para avaliar o desempenho dos agentes. A pontuação é calculada com base no número de unidades eliminadas, no número de unidades sobreviventes e na captura da base inimiga, conforme a interface do jogo apresentada na figura 2.

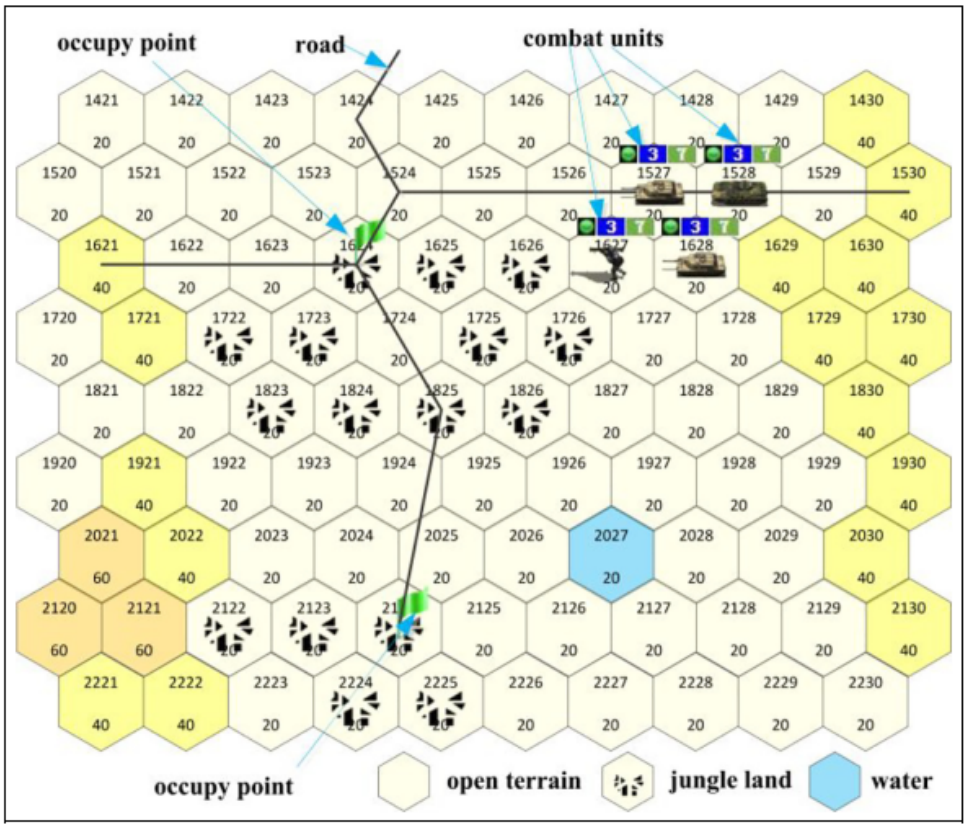

Figura 2: Interface *ArmorCombat*
Fonte: ZHANG e XUE (2020)

Já no artigo de Sun et al. (2020) eles explicam que o jogo é baseado em um sistema de confronto por turnos com modelos de aprendizagem por reforço profunda. Eles projetam um algoritmo Q-learning para alcançar a tomada de decisão inteligente, que é baseado no DQN (*Deep Q Network*) para modelar comportamentos complexos do jogo. Eles também introduzem um algoritmo baseado em conhecimento a priori PK-DQN (*Prior Knowledge-Deep Q Network*) para melhorar o algoritmo DQN. Eles realizam experimentos para demonstrar a eficácia do algoritmo PK-DQN em derrotar o alto nível de oponentes baseados em regras, conforme interface do jogo apresentado na figura 3.

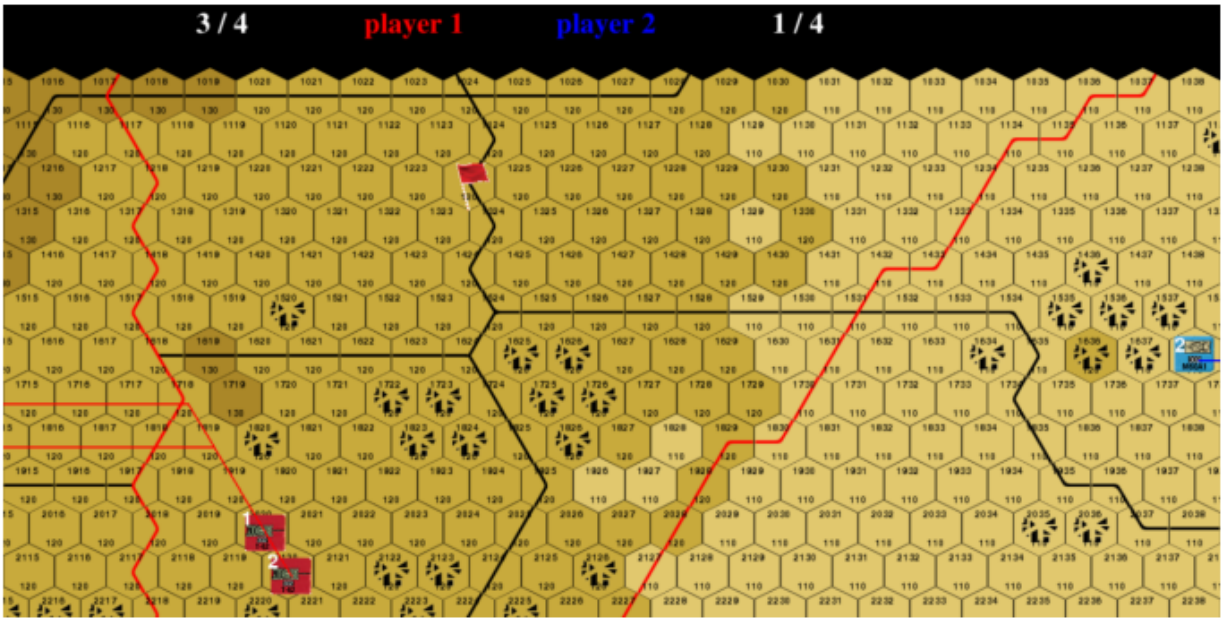

Figura 3: Interface jogo tático de guerra
Fonte: SUN et al. (2020)

Apesar de não descrever com exatidão o ambiente do jogo em que realizou seus testes, o jogo citado por Sun et al. (2020) guarda similaridades com descrito por Zhang e Xue (2020), como os aspectos do terreno, formato da célula, e a base em turnos que propiciam a possibilidade de uma comparação entre os resultados em termos relativos.

4.3 Comparação dos resultados dos artigos

Os dois artigos que foram comparados são: Actor–critic-based decision-making method for the artificial intelligence commander in tactical wargame de Junfeng Zhang e Qing Xue (ZHANG e XUE, 2020) e Research and

Implementation of Intelligent Decision Based on a Priori Knowledge and DQN Algorithms in Wargame Environment de Yuxiang Sun et al. (SUN et al., 2020).

Ambos os artigos propõem métodos baseados em aprendizagem por reforço profundo para modelar o comportamento e a decisão de agentes inteligentes em jogos de guerra táticos. Eles desenvolvem jogos de guerra como ambientes de pesquisa, que contêm scripts de inteligência artificial embutidos e suportam o modo de combate máquina-máquina. Eles realizam experimentos para avaliar o desempenho dos agentes em diferentes cenários e condições.

O primeiro artigo, de Zhang e Xue (2020), propõe um método baseado em ator-crítico para resolver o problema de decisão do comandante de inteligência artificial em um jogo de guerra tático chamado ArmorCombat. O método usa uma rede neural convolucional para representar a situação do campo de batalha e um método de aprendizagem por reforço para testar diferentes estratégias táticas. O objetivo do jogo é eliminar todas as unidades inimigas ou capturar a base inimiga.

O segundo artigo, de Sun et al. (2020), apresenta um algoritmo baseado em conhecimento a priori para melhorar o algoritmo DQN (Deep Q Network), que acelera a velocidade de convergência e a estabilidade do algoritmo. O algoritmo usa um sistema de confronto por turnos com modelos de aprendizagem por reforço profundo. O objetivo do jogo é derrotar o alto nível de oponentes baseados em regras.

Os resultados dos dois artigos podem ser comparados nos seguintes aspectos:

A arquitetura da rede neural: O primeiro artigo usa uma rede neural convolucional com três camadas ocultas, que recebe como entrada uma imagem do mapa do jogo e produz como saída uma distribuição de probabilidade sobre as ações possíveis do comandante (ZHANG e XUE, 2020, p. 5). O segundo artigo usa uma rede neural convolucional com quatro camadas ocultas, que recebe como entrada uma matriz binária que representa o estado do jogo e produz como saída um valor Q para cada ação possível do agente (SUN et al., 2020, p. 4).

O algoritmo de aprendizagem por reforço: O primeiro artigo usa um algoritmo baseado em ator-crítico, que consiste em dois componentes: um ator, que seleciona as ações com base na política aprendida pela rede neural, e um crítico, que avalia as ações com base na função valor estimada pela rede neural (ZHANG e XUE, 2020, p. 5). O segundo artigo usa um algoritmo baseado em DQN, que consiste em

uma rede neural que aproxima a função Q ótima, que representa o valor esperado de cada ação em cada estado (SUN et al., 2020, p. 4).

A incorporação do conhecimento a priori: O primeiro artigo não incorpora nenhum conhecimento a priori no algoritmo de aprendizagem por reforço, mas sim no script de inteligência artificial embutido no jogo, que fornece algumas regras básicas para o comportamento das unidades militares (ZHANG e XUE, 2020, p. 4). O segundo artigo incorpora o conhecimento a priori no algoritmo PK-DQN (Prior Knowledge-Deep Q Network), que modifica o algoritmo DQN para considerar as informações prévias sobre o jogo, como as características das unidades, os tipos de terreno e as estratégias dos oponentes (SUN et al., 2020, p. 5).

A avaliação do desempenho dos agentes: O primeiro artigo avalia o desempenho dos agentes com base na pontuação obtida no jogo, que é calculada com base no número de unidades eliminadas, no número de unidades sobreviventes e na captura da base inimiga (ZHANG e XUE, 2020, p. 5). O segundo artigo avalia o desempenho dos agentes com base na taxa de vitória contra os oponentes baseados em regras, que representam diferentes níveis de dificuldade (SUN et al., 2020, p. 6).

Os resultados dos dois artigos mostram que os agentes baseados em aprendizagem por reforço profundo são capazes de aprender a tomar decisões inteligentes em jogos de guerra táticos, e que os agentes que incorporam o conhecimento a priori têm uma vantagem sobre os agentes que não incorporam. No entanto, os resultados também mostram que os agentes ainda têm limitações e desafios, como a dependência do cenário, a variabilidade dos resultados, a necessidade de ajuste de parâmetros e a complexidade computacional.

Zhang e Xue (2020) realizam um experimento de combate entre um agente baseado em aprendizagem por reforço profundo e um agente baseado em regras (RB) em um cenário de terreno da selva. O experimento é repetido e são retirados dados a cada 50 interações entre os agentes.

Os resultados mostram que o agente baseado em DRL tem uma pontuação média mais alta do que o agente baseado em regras, indicando que o agente baseado em DRL aprende a tomar decisões mais eficazes no jogo de guerra tático. As curvas são mostradas na figura 4:

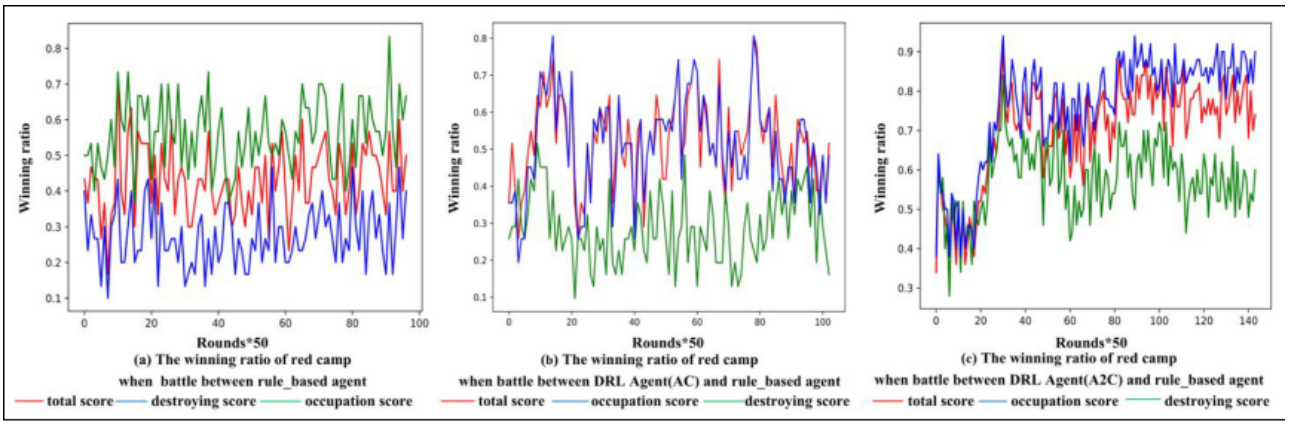

Figura 4: Gráfica da performance de vitórias dos embates entre agente baseado em regras versus baseado em ator crítico e baseado em ator crítico avançado.
Fonte: Zhang e Xue (2020)

As curvas apresentadas na Figura 4, mostram que o agente baseado em DRL melhora gradualmente sua pontuação ao longo do tempo e se estabiliza com média de vitórias superiores a 75%, mas também apresenta algumas flutuações e oscilações. Os autores atribuem essas variações à natureza estocástica do algoritmo de aprendizagem por reforço, à complexidade do ambiente e à aleatoriedade das ações do oponente.

Os autores afirmam que exemplos mostram que o agente baseado em DRL é capaz de explorar as vantagens das unidades militares, como o alcance, a velocidade e o dano, e adaptar-se às mudanças do campo de batalha, como o tipo de terreno e a posição do inimigo. O agente baseado em DRL também é capaz de capturar a base inimiga ou eliminar todas as unidades inimigas, alcançando o objetivo do jogo. Por outro lado, o agente baseado em regras segue um conjunto fixo de regras, que nem sempre são adequadas para o cenário ou para a situação. O agente baseado em regras também tende a ser mais passivo e defensivo, evitando o confronto direto com o inimigo.

Sun et al. (2020) realizaram dois experimentos para avaliar o desempenho do algoritmo PK-DQN (Prior Knowledge-Deep Q Network) em comparação com o algoritmo DQN convencional. O primeiro experimento comparou a velocidade de convergência e a estabilidade dos dois algoritmos, medindo a variação da função custo ao longo dos episódios de treinamento. O segundo experimento compara a

taxa de vitória dos dois algoritmos contra os oponentes baseados em regras conforme a Figura 5:

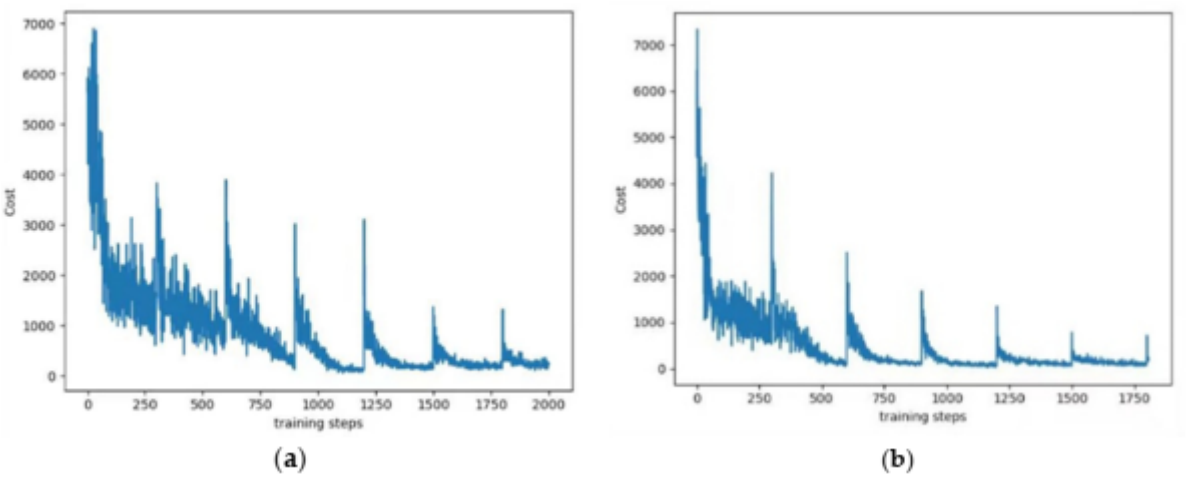

Figura 5: (a) Desempenho da curva de aprendizagem DQN em ambiente de simulação; (b) Desempenho de a curva de aprendizado PK-DQN em ambiente de simulação. A coordenada X representa o número de etapas de treinamento, ou seja, as etapas totais do treinamento de confronto; a coordenada Y representa o tamanho de o valor da curva de custo Li(θi)
Fonte: Sun et al. (2020)

Segundo Sun et al. (2020) em termos de velocidade de convergência, fica clara a maior eficiência do PK-DQN que estabiliza após pouco menos de 1000 passos, ao contrário do DQN que leva aproximadamente 1000 passos.

No que tange à taxa de vitória, o algoritmo PK-DQN se sobressai em relação ao algoritmo DQN, indicando que o algoritmo PK-DQN aprende a tomar decisões mais eficientes no jogo de guerra tático. As curvas são mostradas na figura 6 e 7:

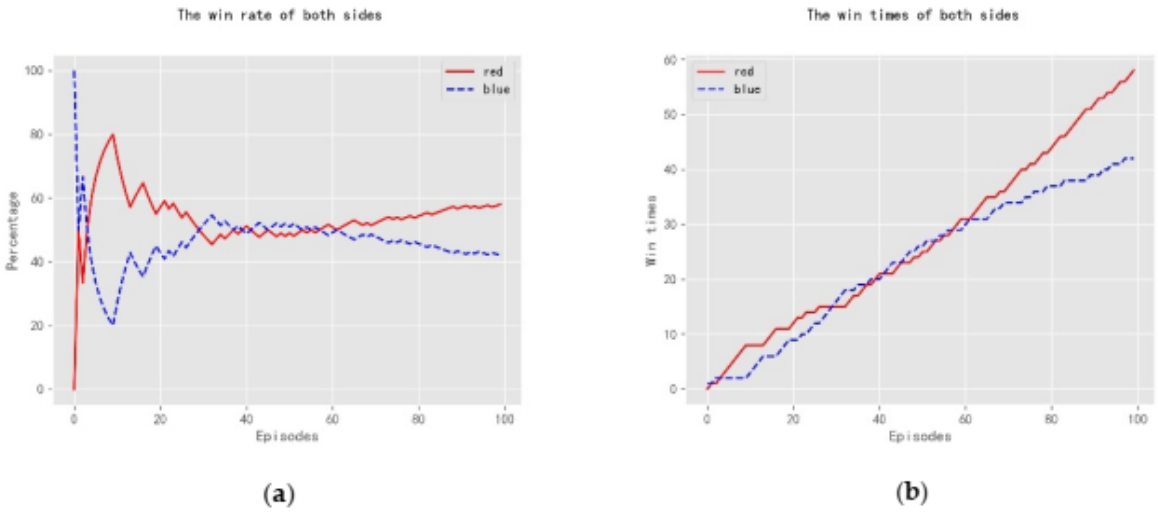

Figura 6: (a) Taxa de vitórias: o lado vermelho é o algoritmo inteligente AI do DQN e o lado azul é IA baseada em regras; (b) Tempos de vitória: o lado vermelho é o algoritmo inteligente AI do DQN e o lado azul é IA baseada em regras; A taxa de vitórias e o número de vitórias para os lados vermelho e azul. a primeira rodada ganha, então um lado começa do 1 e o outro do 0.
Fonte: Sun et al. (2020)

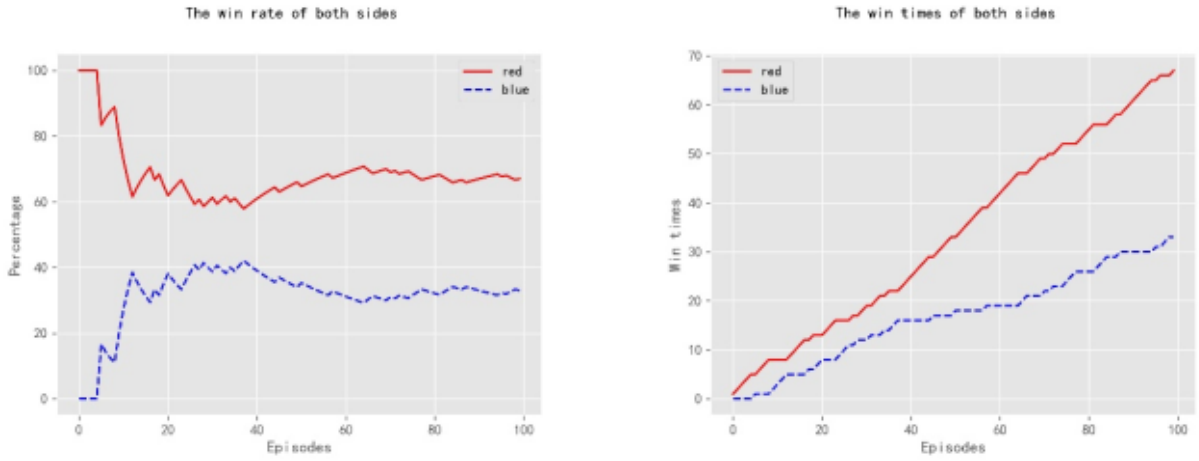

Figura 7: (a) Taxa de vitórias: o lado vermelho é o AI do algoritmo inteligente PK-DQN e o lado azul é IA baseada em regras; (b) Tempos de vitória: o lado vermelho é o algoritmo inteligente AI do PK-DQN e o lado azul é IA baseada em regras; A taxa de vitórias e o número de vitórias para os lados vermelho e azul. a primeira rodada ganha, então um lado começa do 1 e o outro do 0.
Fonte: Sun et al. (2020)

Sun et al. (2020) explicam que o algoritmo PK-DQN apresenta, como esperado, uma taxa de vitória média bem superior ao algoritmo DQN dadas as mesmas quantidades de interações.

5 Conclusão

Os resultados dos dois artigos mostram que os agentes baseados em aprendizagem por reforço profundo são capazes de aprender a tomar decisões

inteligentes em jogos de guerra táticos, e que os agentes que incorporam o conhecimento a priori têm uma vantagem sobre os agentes que não incorporam. No entanto, os resultados também mostram que os agentes ainda têm limitações e desafios, como a dependência do cenário, a variabilidade dos resultados, a necessidade de ajuste de parâmetros e a complexidade computacional.

A implementação do conhecimento a priori parece ser um caminho adequado para o desenvolvimento de agentes que em um futuro possam assessorar comandantes táticos em como agir no campo de batalha, visto que não se deseja a partir do treinamento desses algoritmos criar uma nova doutrina de emprego dos meios mas sim que o agente de IA consiga ler a imensa quantidade de dados disponíveis e enquadrado nas regras do combate dite uma Linha de Ação.

Os dois artigos contribuem para o avanço da pesquisa sobre a tomada de decisão do comandante de inteligência artificial em jogos de guerra táticos, que é um tema relevante e desafiador para a ciência da computação e para as ciências militares. Eles também abrem possibilidades para futuros trabalhos, como a aplicação dos métodos propostos em outros tipos de jogos ou ambientes, a incorporação de outros tipos de conhecimento a priori ou aprendido, a melhoria da arquitetura da rede neural ou do algoritmo de aprendizagem por reforço, e a avaliação dos agentes em cenários mais realistas ou contra adversários humanos.

Referências Bibliográficas

BISHOP, C. M. Pattern Recognition and Machine Learning. New York: Springer, 2006.

BRASIL. Marinha do Brasil. Corpo de Fuzileiros Navais. CGCFN 0-1 Manual Básico dos Grupamentos Operativos de Fuzileiros Navais, 2020.

CALAÇA, Capitão. Ciclo O.O.D.A: os ensinamentos de John Boyd. InfoArmas: O Maior portal sobre armas da América Latina. Disponível em: https://infoarmas.com.br/ciclo-o-o-d-a-os-ensinamentos-de-john-boyd/. Acesso em: 30 out. 2022.

EXÉRCITO DOS ESTADOS UNIDOS. FM 3-0 Operations. Washington, D.C.: Departamento de Defesa dos Estados Unidos, 2020.

FEICKERT, A. Defense Primer: Army Multi Domain Operations (MDO). Washington, D.C: Congressional Research SVC, 2021.

GOECKS, V. G.; WAYTOWICH, N.; WATKINS, D.; PRAKASH, B. Combining learning from human feedback and knowledge engineering to solve hierarchical tasks in minecraft. arXiv Disponível em: https://doi.org/10.48550/arXiv.2112.03482 (accessed on December 7,2021).

HAMMOND, Grant T. The Mind of War: John Boyd and American Security. Washington: Smithsonian Books, 2001.

KASE SE; HUNG CP; KRAYZMAN T; HARE JZ; RINDERSPACHER BC; SU SM. The Future of Collaborative Human-Artificial Intelligence Decision-Making for Mission Planning. Frontiers in Psychology, v. 13, p. 850628, 2022.

KURZWEIL, R. The Singularity is Near: When Humans Transcend Biology. New York: Viking Press, 2005.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep Learning. Nature, v. 521, n. 7553, p. 436-444, 2015.

MITCHELL, T. M. Machine Learning. New York: McGraw Hill, 1997.

MNINH, V.; KAVUKCUOGLU, K.; SILVER, D.; et al. Human-level control through deep reinforcement learning. Nature, v. 518, n. 7540, p. 529-533, 2015.

NILSSON, N. J. Artificial Intelligence: A New Synthesis. San Francisco: Morgan Kaufmann Publishers Inc., 1998.

RUSSEL, S.; NORVIG, P. Inteligência Artificial. 4 ed., São Paulo: Pearson Prentice Hall Brasil Ltda., 2022.

SCHUMPETER, Joseph A. Capitalismo, Socialismo e Democracia. Rio de Janeiro: Editora Fundo de Cultura Ltda., 1942.

SILVA, A. et al. Processo de Decisão de Markov — Parte 1 - Turing Talks. Disponível em:https://medium.com/turing-talks/aprendizado-por-refor%C3%A7o-2-processo-de-decis%C3%A3o-de-markov-mdp-parte-1-84e69e05f007

SILVA, R et al. Aprendizado por Reforço Profundo: Uma Introdução. Disponível em: https://medium.com/data-hackers/aprendizado-por-refor%C3%A7o-profundo-uma-introdu%C3%A7%C3%A3o-aa9c0f8dcbab

SUTTON, R. S.; BARTO, A. G. Reinforcement Learning: An Introduction. 2nd ed., Cambridge, MA: MIT Press, 2018.

SUN, Y.; LIU, J.; ZHANG, J.; et al. Research and Implementation of Intelligent Decision Based on a Priori Knowledge and DQN Algorithms in Wargame Environment. IEEE Access, v. 8, p. 218329-218339, 2020.

VINYALS, O.; BABUSCHKIN, I.; CZARNECKI, W. M.; et al. Grandmaster level in StarCraft II using multiagent reinforcement learning. Nature, v. 575, n. 7782, p. 350-354, 2019.

ZAWISLAK, P. A.; ALVES, A. C.; TELLO-GAMARRA, J.; BARBIEUX, D.; REICHERT, F. M. Innovation Capability: From Technology Development to Transaction Capability. Bingley: Emerald Group Publishing Limited, 2012.

ZHANG J; XUE Q. Actor–critic-based decision-making method for the artificial intelligence commander in tactical wargame. Neural Computing and Applications, v. 32, n. 16, p. 12169-12181, 2020.