PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
PRÓ-REITORIA DE GRADUAÇÃO
ESCOLA POLITÉCNICA E DE ARTES
GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

O IMPACTO DA APRENDIZAGEM POR REFORÇO EM JOGOS COMPLEXOS

Paulo Rios de Castro Pontes

GOIÂNIA-GO,

2024

Paulo Rios de Castro Pontes

O IMPACTO DA APRENDIZAGEM POR REFORÇO EM JOGOS COMPLEXOS

Trabalho de Conclusão de Curso apresentado na escola Politécnica e de artes, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação.

Orientado(a): Paulo Rios de Castro Pontes

Orientador(a): Gustavo Siqueira Vinhal.

Resumo

Este trabalho explora a aplicação da Aprendizagem por Reforço (RL) em jogos complexos, com foco em sua contribuição para tornar esses jogos mais desafiadores, envolventes e inovadores. O RL é uma técnica central no desenvolvimento de agentes inteligentes que aprendem a tomar decisões baseadas em recompensas e punições. Em jogos como Go, Dota 2 e StarCraft II, o RL tem permitido que os agentes melhorem suas estratégias, superando até mesmo jogadores humanos em ambientes altamente desafiadores. O futuro do RL em jogos também é explorado, com ênfase em questões éticas, transparência e a necessidade de algoritmos mais eficientes. A conclusão destaca a importância do RL como uma ferramenta inovadora para o desenvolvimento de jogos mais desafiadores e a transferência dessa tecnologia para outras áreas, como robótica e saúde. O trabalho sugere que futuros estudos devem focar em aprimorar a eficiência computacional, generalização dos agentes e a integração do RL com outras abordagens de IA, além de considerar as implicações éticas e transparentes no desenvolvimento desses sistemas.

Palavras-chave: Inteligência artificial, Aprendizagem por reforço, jogos complexos, redes neurais.

Abstract

This paper explores the application of Reinforcement Learning (RL) in complex games, focusing on its contribution to making these games more challenging, engaging, and innovative. RL is a central technique in the development of intelligent agents that learn to make decisions based on rewards and punishments. In games such as Go, Dota 2, and StarCraft II, RL has allowed agents to improve their strategies, even outperforming human players in highly challenging environments. The future of RL in games is also explored, with emphasis on ethical issues, transparency, and the need for more efficient algorithms. The conclusion highlights the importance of RL as an innovative tool for the development of more challenging games and the transfer of this technology to other areas, such as robotics and healthcare. The paper suggests that future studies should focus on improving computational efficiency, agent generalization, and the integration of RL with other AI approaches, in addition to considering the ethical and transparency implications of developing these systems.

Keywords: Artificial intelligence, Reinforcement learning, complex games, neural networks.

LISTA DE FIGURAS

Figura 1: Sistema de recompensa 13
Figura 2: Tabela de recompensa 14
Figura 3: Rede Neural de ações 17
Figura 4: Imagem jogo GO 20
Figura 5: Imagem jogo Dota 2 21
Figura 6: Imagem jogo Starcraft 2 22
Figura 7: Complexidade da aprendizagem por reforço 22

LISTA DE SIGLAS

IA – INTELIGENCIA ARTIFICIAL

ML- MACHINE LEARNING

RL - Aprendizagem por Reforço (Reinforcement Learning)

DQN - Deep Q-Network

Dueling DQN - Dueling Deep Q-Network

OpenAI - Open Artificial Intelligence

AlphaGo - Programa de IA desenvolvido para jogar Go

AlphaStar - Programa de IA desenvolvido para jogar StarCraft II

Meta-RL - Meta-Reinforcement Learning

Transfer Learning - Aprendizado por Transferência

GPT - Generative Pre-trained Transformer

SUMARIO

1 INTRODUÇÃO 8
1.1 OBJETIVOS 10
1.1.1 OBJETIVOS GERAIS 10
1.1.2 OBJETIVOS ESPECÍFICOS 10
1.2 METODOLOGIA 10
2 FUNDAMENTOS DA APRENDIZAGEM POR REFORÇO 12
2.1 CONCEITOS FUNDAMENTAIS 12
2.2 ALGORITIMOS DE APRENDIZAGEM POR REFORÇO 13
2.2.1 MÉTODOS BASEADOS EM VALOR 13
2.2.2 MÉTODOS BASEADOS EM POLÍTICA 13
2.2.3 METODOS HÍBRIDOS 13
2.3 EVOLUÇÃO DO R1 E O USO DE REDES NEURAIS 13
2.4 DESAFIOS E LIMITAÇÕES 14
2.5 RL EM JOGOS COMPLEXOS 14
3 APLICAÇÃO DA APRENDIZAGEM POR REFORÇO EM JOGO COMPLEXO .. 14
3.1 JOGOS COM LABORATÓRIOS DE APRENDIZADO 15
3.2 ALPHAGO E O AVANÇO NO JOGO DE GO 15
3.3 OPEN AI FIVE E O DESAFIO DOS JOGOS MULTIAGENTES 15
3.4 STARCRAFT II E A TOMADA DE DECISÃO EM ABIENTE INCERTO 16
3.5 BENEFICIOS ALÉM DO ENTRETERIMENTO 16
3.6 DESAFIOS E PERSPECTIVAS 17
4 DESAFIOS E AVANÇOS RECENTES NA APLICAÇÃO DA AR 17
4.1 ESCABILIDADE E ESPAÇOS DE ESTADO DE ALTA DIMENSÃO 17
4.2 EXPLORAÇÃO 18
4.3 ROBUSTEZ EM AMBIESTES INCERTOS 19
4.4 REDUÇÃO DE CUSTOS COMPUTACIONAIS 19
4.5 TENDENCIAS EMERGENTES NO RL 19
4.5.1 RL MULTIAGENTES 19
4.5.2 META-RL E TRANSFER LEARNING 20
4.5.3 RL COM CONSCIENTIZAÇÃO ÉTICA 20
4.5.4 INTEGRAÇÃO COM MODELOS FUNDAMENTAIS 20

4.6 IMPACTOS FUTURÍSTICOS.....21
5 CONCLUSÃO 22
5.1 TRABALHOS FUTUROS.....22
REFERENCIAS.....23

1 Introdução

A aprendizagem por reforço (*Reinforcement Learning – RL*) é uma abordagem inovadora da inteligência artificial (IA) que tem revolucionado diversos setores da tecnologia, desde a robótica até a análise de dados financeiros. Esse método de aprendizado baseia-se em um conceito simples, porém poderoso: um agente interage com um ambiente, toma decisões e aprende com base nas recompensas ou penalidades recebidas. Diferentemente do aprendizado supervisionado, que depende de um grande volume de dados rotulados, o RL explora e experimenta soluções de forma autônoma para descobrir as melhores estratégias de ação em ambientes dinâmicos e incertos (SUTTON; BARTO, 2018).

Nos últimos anos, o RL consolidou-se como uma ferramenta essencial no desenvolvimento de sistemas capazes de lidar com problemas complexos. Um dos campos mais notáveis de sua aplicação é o dos jogos, particularmente aqueles que apresentam alta complexidade estratégica e operacional, como xadrez, Go, Dota 2 e StarCraft II. Esses jogos, além de oferecerem desafios significativos, servem como laboratórios para testar e refinar algoritmos que podem ser aplicados em outros contextos do mundo real. Historicamente, os programas de jogos utilizavam regras fixas ou análises estatísticas predefinidas, mas com o advento do RL, tornou-se possível criar agentes que aprendem a partir de tentativa e erro, adaptando-se dinamicamente e desenvolvendo novas estratégias (GOODFELLOW; BENGIO; COURVILLE, 2016).

Exemplos de sucesso demonstram a eficácia do RL em jogos complexos. Em 2016, o AlphaGo, desenvolvido pela DeepMind, superou o campeão mundial de Go, Lee Sedol, em uma série histórica de partidas. Utilizando redes neurais profundas em combinação com métodos de busca em árvores, o AlphaGo não apenas aprendeu estratégias a partir de jogos humanos, mas também foi capaz de superar esses padrões, criando abordagens inéditas (SILVER et al., 2016). Outro exemplo impressionante foi o OpenAI Five, uma equipe de agentes treinada por meio de RL para jogar Dota 2. Em 2022, o OpenAI Five venceu jogadores profissionais no Campeonato Mundial de Dota 2, demonstrando a capacidade dos agentes de lidar com ambientes complexos e multiagentes em tempo real (OPENAI, 2023).

Além do impacto técnico, o setor de jogos é relevante pelo seu enorme peso econômico. Em 2021, o mercado global de jogos movimentou aproximadamente US$

192 bilhões. O crescimento contínuo dessa indústria torna os jogos não apenas um campo de estudo para a IA, mas também um motor econômico importante. Investir em tecnologias como o RL no contexto dos jogos pode criar experiências mais imersivas e desafiadoras, ao mesmo tempo em que avança a tecnologia para aplicações fora do entretenimento, como saúde, transporte e automação industrial (MARTINS, 2018).

A interseção entre RL e jogos complexos também é de grande interesse acadêmico e prático, pois os desafios enfrentados nesses ambientes se assemelham a problemas do mundo real. Jogos como StarCraft II, por exemplo, envolvem tomada de decisão em tempo real, incerteza nas informações disponíveis e a necessidade de coordenação entre múltiplos agentes. Essas características são similares às encontradas em situações como planejamento logístico, gerenciamento de sistemas autônomos e controle de tráfego aéreo (VINYALS et al., 2019). Assim, avanços no RL aplicados a jogos podem se traduzir em melhorias diretas para diversas áreas.

Outro aspecto que torna a investigação dessa área tão relevante é sua capacidade de impulsionar a inovação na criação de jogos. A aprendizagem por reforço pode ser utilizada não apenas para treinar agentes competitivos, mas também para projetar sistemas mais adaptativos que ajustem a dificuldade ou ofereçam experiências personalizadas aos jogadores, aumentando o engajamento. Silver (2017) argumenta que o RL pode transformar a maneira como jogos são projetados, permitindo a criação de mundos virtuais dinâmicos e interativos que se adaptam ao comportamento do jogador, ampliando os limites do entretenimento digital.

Portanto, este trabalho tem como objetivo explorar o papel do RL na criação de jogos complexos mais desafiadores, inovadores e envolventes, destacando os principais avanços na área, como o uso de redes neurais profundas, e os desafios técnicos, como a escalabilidade e o alto custo computacional. Além disso, este estudo busca responder à seguinte questão: **Como a aprendizagem por reforço pode ser usada para aprimorar jogos complexos mais desafiadores, envolventes e inovadores?**

1.1 Objetivos

1.1.1 Objetivo Geral:

Investigar e analisar o uso do Aprendizado por Reforço (RL) como uma abordagem para tornar os jogos mais desafiadores, envolventes e inovadores.

1.1.2 Objetivos Específicos:

- Realizar a Revisão Bibliográfica
- Avaliar o impacto da aprendizagem por reforços nos jogos complexos
- Identificar desafios da aprendizagem por reforço
- Identificar vantagens da aprendizagem por reforço para jogos

1.2 Metodologia:

Esta pesquisa segundo sua natureza é um resumo de assunto. Fundamenta-se em trabalhos da área de conhecimento do projeto, buscando reunir, analisar, avaliar e discutir conhecimentos e informações (Wazlawick, 2014).

Segundo os objetivos é uma pesquisa exploratória. A pesquisa exploratória é aquela em que o autor não tem necessariamente uma hipótese ou objetivo definido em mente. Ela pode ser considerada, muitas vezes, como o primeiro estágio de um processo de pesquisa mais longo. Na pesquisa exploratória, o autor vai examinar um conjunto de fenômenos, buscando anomalias que não sejam ainda conhecidas e que possam ser, então, a base para uma pesquisa mais elaborada. Segundo aos procedimentos técnicos, esta pesquisa é bibliográfica e experimental. (WAZLAWICK, 2014).

Estuda-se artigo, teses, livros entre outras publicações. Tal pesquisa é necessária, pois é com ela que o pesquisador vai fundamentar-se para prosseguir com sua pesquisa. Além disso, o pesquisador realiza experimentos manipulando variáveis da realidade a fim de observar e analisar os efeitos obtidos por tal alteração (WAZLAWICK, 2014).

De acordo com Gil (2017), a pesquisa bibliográfica se desenvolve a partir das seguintes etapas:

a) Escolha do tema: O impacto da aprendizagem por reforço em jogos complexos

b) Revisão preliminar da literatura: Fazer o levantamento bibliográfico preliminar de periódicos e artigos relacionados ao assunto de pesquisa. No caso, sobre a aplicação de aprendizagem por reforço em jogos - foram feitas na base de dados da CAPES, repositório da PUCGO e Google Acadêmico.

c) Formulação da pergunta de pesquisa: **Como a aprendizagem por reforço pode ser usada para criar jogos complexos mais desafiadores, envolventes e inovadores?**

d) Identificação das fontes: Identificar as fontes bibliográficas que forneçam informações relevantes para responder ao problema proposto, consultando dissertações, periódicos científicos, obras de referência, entre outros.

e) Análise do conteúdo: Examinar as informações e dados coletados das fontes selecionadas, relacionando-os à pergunta de pesquisa e avaliando a consistência dos argumentos apresentados.

f) Fichamento: O fichamento foi realizado com ênfase em trabalhos relevantes para a pesquisa. No entanto, documentos com contribuições médias ou baixas também foram analisados para enriquecer as reflexões do autor. Essa abordagem permitiu que o TCC fosse escrito com uma base sólida.

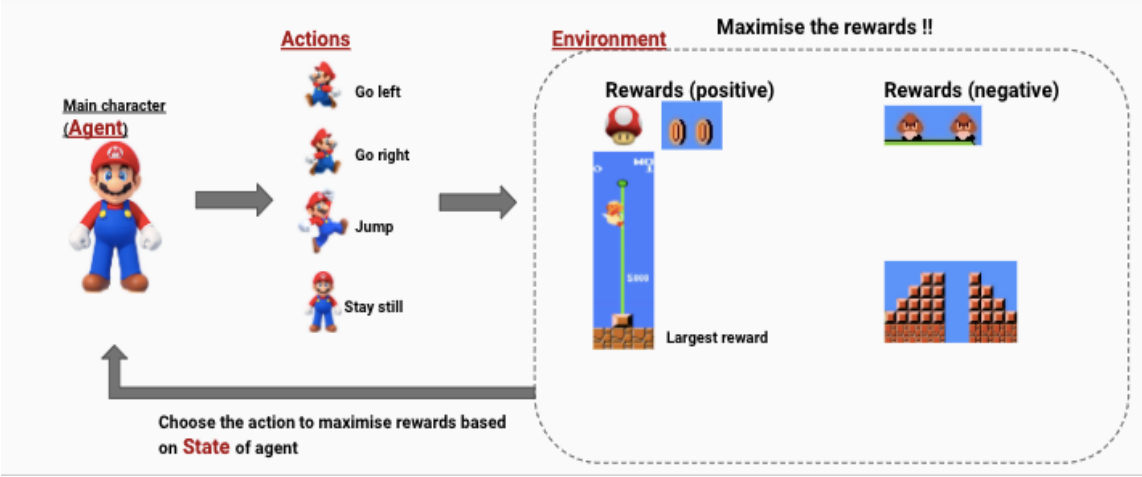
g) Escrita do Trabalho de Conclusão de Curso.

Quanto aos procedimentos técnicos a pesquisa experimental implica ter uma ou mais variáveis experimentais que podem ser controladas pelo pesquisador (o fato de usar ou não determinada técnica, por exemplo), e uma ou mais variáveis observadas, cuja medição poderá levar, possivelmente, à conclusão de que existe algum tipo de dependência com a variável experimental (por exemplo, avaliar a produtividade dos programadores em pontos de função por dia de trabalho e verificar se a técnica aumenta significativamente esse valor).

2 Fundamentos da Aprendizagem por Reforço

A aprendizagem por reforço (*Reinforcement Learning – RL*) é um ramo do aprendizado de máquina baseado na interação entre agentes e ambientes. O objetivo principal do RL é capacitar agentes a tomarem decisões inteligentes em cenários complexos, utilizando um processo contínuo de tentativa, erro e aprendizado baseado em recompensas e penalidades (SUTTON; BARTO, 2018). Na figura 1 é possível observar esse sistema de recompensa, com seus estados, ações e recompensas.

Figura 1: Sistema de recompensa



Fonte: Medium, 2024

Conforme apresentado na Figura 1, observa-se que o ambiente desempenha um papel significativo no sistema de recompensas, no qual itens são caracterizados como estímulos positivos, enquanto os inimigos representam estímulos negativos

2.1 Conceitos Fundamentais

O RL opera com base em um ciclo de interação entre um agente (o sistema que aprende) e um ambiente (o cenário em que o agente atua). Em cada iteração, o agente observa o estado atual do ambiente, realiza uma ação e recebe um feedback em forma de recompensa. Esse feedback é utilizado para ajustar as ações futuras, de forma a maximizar as recompensas acumuladas ao longo do tempo (SUTTON; BARTO, 2018).

Os elementos centrais do RL incluem:

- Estado (s): Representação da situação atual do ambiente percebida pelo agente.
- Ação (a): Decisão tomada pelo agente com base no estado atual.
- Recompensa (r): Feedback numérico recebido pelo agente após realizar uma ação.
- Política (π): Estratégia que mapeia estados para ações.
- Função de Valor (V): Mede a qualidade de um estado, indicando a recompensa esperada a partir dele.
- Função Q (Q): Mede a qualidade de uma ação em um estado específico.

Esses componentes trabalham juntos para guiar o aprendizado do agente, permitindo que ele descubra a melhor política para alcançar seus objetivos.

Com esses componentes é possível criar uma tabela de recompensa para cada ação demonstrando as ações com mais valores como na figura 2.

Figura 2: Tabela de Recompensa

	000 100	000 010	000 001	100 000	010 000	001 000
↑	0.2	0.3	1.0	-0.22	-0.3	0.0
↓	-0.5	-0.4	-0.2	-0.04	-0.02	0.0
→	0.21	0.4	-0.3	0.5	1.0	0.0
←	-0.6	-0.1	-0.1	-0.31	-0.01	0.0

Fonte: deeplearningbook, 2024

A Figura 2 apresenta um exemplo de uma Tabela de recompensa, utilizada no contexto do Aprendizado por Reforço, para a tomada de decisão em um ambiente de jogo. Nessa tabela, cada célula representa um valor Q associado a uma determinada ação, que indica a qualidade ou a expectativa de recompensa ao executar essa ação em um estado específico. Os valores Q são atualizados com

base na equação de Q-Learning, que considera as recompensas imediatas e futuras, com o fator de desconto γ (neste caso, 0,95).

No tabuleiro exibido à esquerda, a posição atual do agente (o carro vermelho) é identificada pelo estado 010 (o segundo estado na tabela). Para esse estado, os valores Q associados às ações possíveis (cima, baixo, direita e esquerda) são fornecidos na tabela:

- **Cima (↑):** -0,3
- **Baixo (↓):** -0,02
- **Direita (→):** 1,0
- **Esquerda (←):** -0,01

Com base nesses valores Q, observa-se que a ação "Direita" possui o maior valor Q (1,0), indicando que ela é a melhor ação a ser tomada no momento, pois proporciona a maior expectativa de recompensa futura. Por outro lado, ações como "Cima" (-0,3) e "Esquerda" (-0,01) possuem valores negativos, sugerindo que essas ações levam a estados menos desejáveis ou recompensas menores.

Portanto, a Tabela de Recompensa serve como uma representação do conhecimento adquirido pelo agente durante o processo de aprendizado, guiando-o para tomar decisões ótimas no ambiente, com base na maximização dos valores Q

2.2 Algoritmos de Aprendizagem por Reforço

Os algoritmos de RL são classificados em três categorias principais: métodos baseados em valor, métodos baseados em política e métodos híbridos.

2.2.1 Métodos Baseados em Valor

Esses algoritmos concentram-se na estimativa da função de valor ou da função Q para determinar as melhores ações a serem tomadas. Um exemplo clássico é o Q-Learning, que busca aprender a função Q ótima iterativamente. Apesar de eficiente em muitos cenários, sua aplicação em ambientes de grande escala é limitada pela alta dimensionalidade dos estados e ações (SUTTON; BARTO, 2018).

2.2.2 Métodos Baseados em Política

Nessa abordagem, o agente aprende diretamente uma política que mapeia estados para ações. Métodos baseados em política, como Policy Gradient, são particularmente úteis em ambientes com grandes espaços de ação ou onde as ações são contínuas (GOODFELLOW; BENGIO; COURVILLE, 2016).

2.2.3 Métodos Híbridos

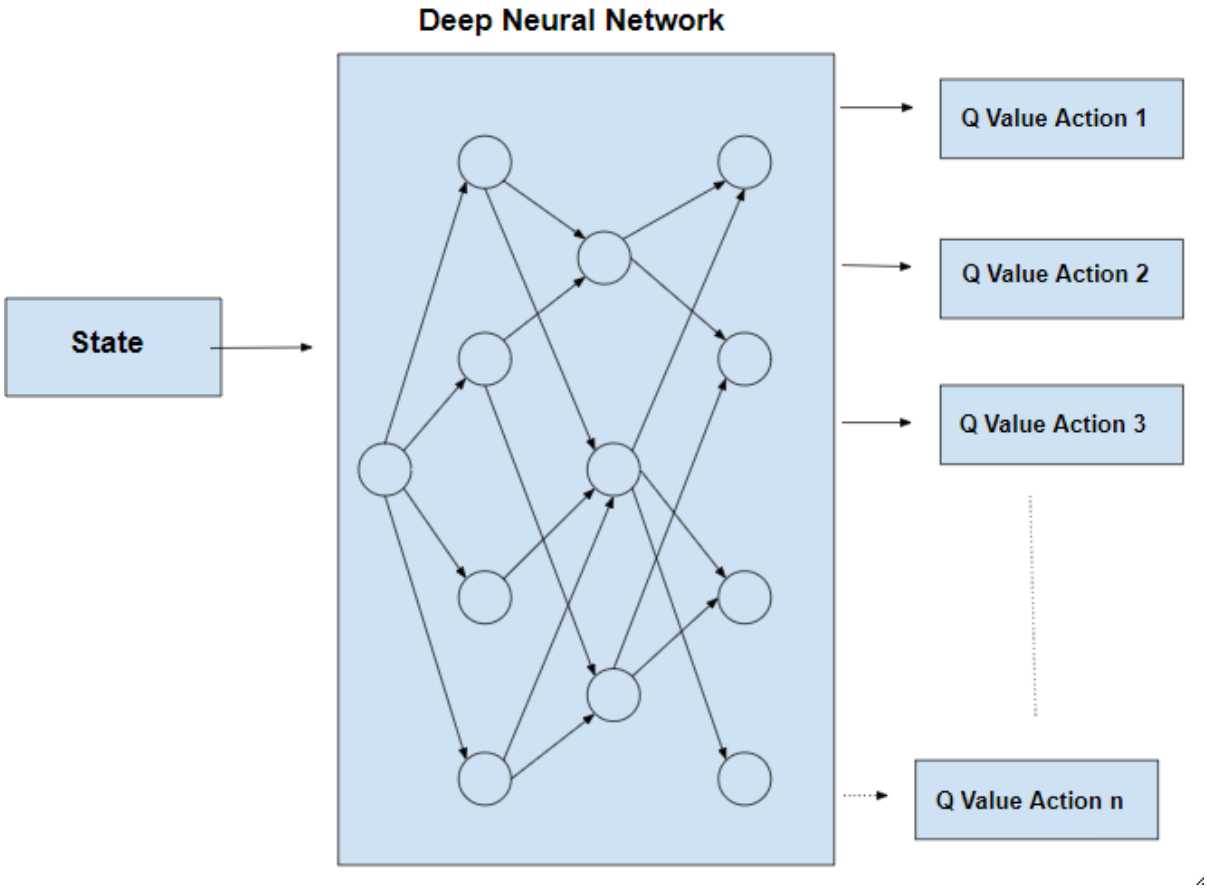
Esses algoritmos combinam características de métodos baseados em valor e política. Um exemplo amplamente utilizado é o Actor-Critic, que utiliza um ator para atualizar a política e um crítico para estimar o valor da política atual (MNIH et al., 2015).

2.3 Evolução do RL e o Uso de Redes Neurais

Com a evolução do aprendizado profundo (deep learning), o RL passou por um salto significativo em suas capacidades. A integração de redes neurais profundas permitiu que algoritmos de RL lidassem com espaços de estado complexos e de alta dimensionalidade. Essa combinação, conhecida como Deep Reinforcement Learning (DRL), tem sido responsável por avanços notáveis, como o desenvolvimento de agentes que superam humanos em jogos complexos (SILVER et al., 2016).

O trabalho de Mnih et al. (2015), que introduziu a técnica Deep Q-Networks (DQN), foi um marco no campo. Usando redes neurais convolucionais para aproximar a função Q, o DQN demonstrou desempenho humano em jogos da plataforma Atari, estabelecendo as bases para a aplicação do DRL em problemas mais desafiadores. Na figura 3 demonstramos as redes neurais de uma aplicação de Deep Reinforcement Learning.

Figura 3: Rede Neural de ações



Fonte: deeplearningbook, 2024

Conforme apresentado na Figura 3, é possível observar a complexidade de uma rede neural de ações voltada para o aprendizado por reforço, na qual diferentes estados podem resultar em ações distintas.

2.4 Desafios e Limitações

Embora o RL tenha alcançado avanços impressionantes, ainda enfrenta desafios significativos:

- Escalabilidade: Ambientes com espaços de estado e ação muito grandes requerem alta capacidade computacional e algoritmos mais eficientes (GOODFELLOW; BENGIO; COURVILLE, 2016).

- Exploração: Encontrar o equilíbrio ideal entre explorar novas ações e explorar as ações conhecidas como melhores continua sendo um problema central no RL (SUTTON; BARTO, 2018).
- Ambientes Incertos: Em ambientes com alta incerteza, a estabilidade e eficiência dos algoritmos podem ser comprometidas (VINYALS et al., 2019).

2.5 RL em Jogos Complexos

Os jogos têm sido amplamente utilizados como campo de testes para RL devido à sua complexidade e desafios dinâmicos. Desde o sucesso do AlphaGo no jogo de Go até o OpenAI Five no Dota 2, o RL provou ser capaz de lidar com cenários estratégicos e imprevisíveis (SILVER et al., 2016; OPENAI, 2023). Além disso, esses avanços têm impulsionado não apenas a criação de jogos mais sofisticados, mas também a aplicação do RL em áreas como logística, saúde e robótica autônoma.

3 Aplicações da Aprendizagem por Reforço em Jogos Complexos

A aprendizagem por reforço tem demonstrado um impacto significativo no desenvolvimento de agentes capazes de enfrentar os desafios impostos por jogos complexos. Esse campo de pesquisa não apenas impulsionou avanços na inteligência artificial, mas também gerou inovações com aplicações práticas fora do domínio do entretenimento. Este capítulo explora os principais casos de uso do RL em jogos, abordando tanto exemplos históricos quanto os avanços mais recentes.

3.1 Jogos como Laboratórios de Aprendizado

Os jogos oferecem ambientes controlados e bem definidos, tornando-os plataformas ideais para testar e refinar algoritmos de RL. Características como regras explícitas, objetivos claros e interatividade permitem que pesquisadores avaliem a eficácia de novas abordagens de maneira mensurável e replicável (SUTTON; BARTO, 2018).

Além disso, jogos complexos, como xadrez, Go e Dota 2, apresentam desafios que simulam problemas do mundo real. Eles envolvem tomada de decisão em tempo real, incertezas, adversários com estratégias dinâmicas e a necessidade de planejamento de longo prazo. Esses aspectos tornam os jogos um microcosmo para o estudo de problemas maiores, como logística, otimização de recursos e controle autônomo (COSTA, 2019).

3.2 AlphaGo e o Avanço no Jogo de Go

O AlphaGo, desenvolvido pela DeepMind, foi um marco na aplicação do RL em jogos. Utilizando uma combinação de redes neurais profundas e métodos de busca em árvores, o AlphaGo conseguiu aprender estratégias a partir de partidas humanas e, posteriormente, superar esses padrões ao jogar contra si mesmo.

Em 2016, o AlphaGo derrotou o campeão mundial de Go, Lee Sedol, em uma série de cinco partidas, marcando a primeira vez que um agente de IA venceu um jogador humano de elite em um jogo tão complexo. A inovação central do AlphaGo foi a introdução de redes neurais que avaliavam posições no tabuleiro e previam

movimentos, combinadas com o uso de RL para aperfeiçoar suas estratégias (SILVER et al., 2016).

Na figura 5 há um exemplo do jogo GO.

Figura 5: Imagem jogo GO

Fonte: baduk, 2024

3.3 OpenAI Five e o Desafio dos Jogos Multiagentes

Outro exemplo notável é o OpenAI Five, uma equipe de agentes de IA treinada para jogar Dota 2. Diferente do Go, o Dota 2 é um jogo multiplayer em tempo real que envolve coordenação entre agentes, tomada de decisão rápida e adaptação a adversários imprevisíveis.

O OpenAI Five utilizou um método de RL chamado **Proximal Policy Optimization (PPO)**, que permite treinar agentes em larga escala de forma eficiente. O treinamento foi realizado em simulações que reproduziam 180 anos de jogo por dia, permitindo que os agentes aprendessem estratégias sofisticadas e colaborativas. Em 2022, o OpenAI Five venceu jogadores profissionais no Campeonato Mundial de Dota 2, demonstrando a capacidade do RL em lidar com ambientes multiagentes altamente complexos (OPENAI, 2023).

Na Figura 6 há um exemplo do jogo Dota 2.

Figura 6: Imagem jogo Dota 2

Fonte: gamearena, 2024

3.4 StarCraft II e a Tomada de Decisão em Ambientes Incertos

StarCraft II, desenvolvido pela Blizzard Entertainment, é amplamente reconhecido como um dos jogos mais desafiadores para agentes de IA devido à sua complexidade estratégica, informações parciais e necessidade de planejamento em tempo real. A DeepMind desenvolveu o **AlphaStar**, um agente que utiliza RL e aprendizado profundo para competir em nível de grande mestre no jogo.

O AlphaStar foi treinado usando uma combinação de aprendizado supervisionado (a partir de dados de jogos humanos) e RL, que permitiu ao agente explorar estratégias inéditas e adaptar-se a diferentes estilos de jogo. Em 2019, o AlphaStar venceu jogadores profissionais de StarCraft II, destacando-se pela sua habilidade de planejar e executar estratégias em um ambiente incerto e dinâmico (VINYALS et al., 2019).

Na figura 7 há um exemplo do jogo Starcraft 2

Figura 7: Imagem jogo Starcraft 2

Fonte: newsBlizzard, 2024

3.5 Benefícios Além do Entretenimento

Embora a aplicação do RL em jogos tenha foco inicial no entretenimento, os avanços gerados por esses estudos têm implicações práticas em diversos setores. A capacidade de tomada de decisão em ambientes dinâmicos, otimização de recursos e coordenação entre múltiplos agentes é valiosa para áreas como:

- **Logística:** Planejamento de rotas e alocação de recursos.
- **Saúde:** Desenvolvimento de sistemas de suporte à decisão em diagnósticos e tratamentos.
- **Robótica:** Controle autônomo em veículos e drones.

Jogos complexos, como StarCraft II, são usados como modelos para resolver problemas como a coordenação de múltiplos agentes em tarefas colaborativas, que têm aplicações diretas em sistemas de transporte autônomo e na indústria 4.0 (COSTA, 2019).

3.6 Desafios e Perspectivas

Apesar dos avanços, o uso do RL em jogos complexos enfrenta desafios significativos:

- **Altos custos computacionais:** O treinamento de agentes, como o AlphaGo e o AlphaStar, requer infraestruturas computacionais robustas, muitas vezes

inacessíveis para pequenos laboratórios (GOODFELLOW; BENGIO; COURVILLE, 2016).

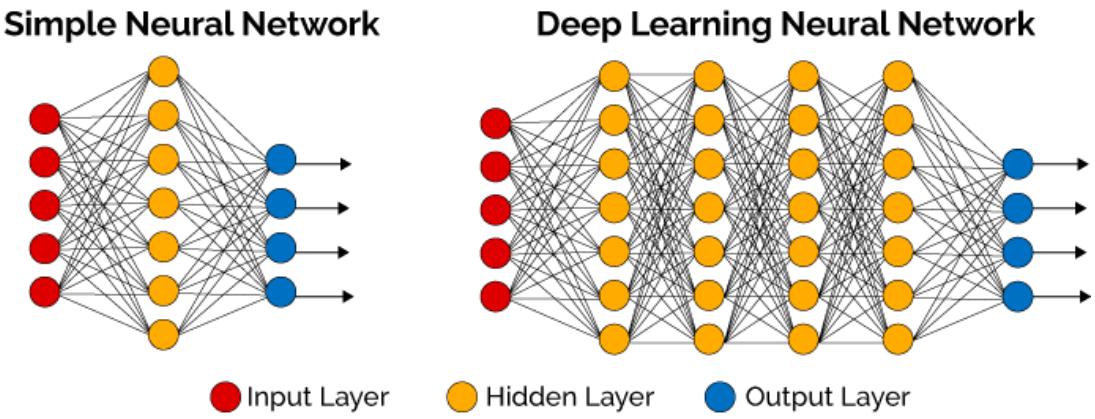
- **Generalização:** Embora os agentes demonstrem alta competência em ambientes específicos, sua transferência para outros domínios ainda é limitada (VINYALS et al., 2019).
- **Ética e transparência:** A crescente capacidade de agentes autônomos levanta questões éticas, como a transparência nos processos de decisão e o impacto social das tecnologias desenvolvidas.

O avanço contínuo da aprendizagem por reforço, combinado com técnicas de aprendizado profundo, promete não apenas transformar a indústria de jogos, mas também influenciar outras áreas críticas. Pesquisas futuras devem focar na redução de custos computacionais, na melhoria da generalização dos agentes e no uso responsável da tecnologia.

4 Desafios e Avanços Recentes na Aplicação da Aprendizagem por Reforço

Apesar do sucesso em jogos complexos, a aprendizagem por reforço (RL) ainda enfrenta barreiras que limitam sua aplicação e expansão. Este capítulo analisa os desafios que permeiam o uso do RL, discute as soluções propostas recentemente e explora as tendências emergentes, com foco em como essas inovações podem melhorar o desempenho de agentes em cenários de jogos complexos e em problemas do mundo real. A figura 8 demonstra a complexidade de RL em comparação a uma rede neural simples.

Figura 8: Complexidade da aprendizagem por reforço



Fonte: Institutodeengenharia, 2024

Na Figura 3, é feita uma comparação entre uma rede neural padrão e uma rede neural baseada em Aprendizado por Reforço (RL). Observa-se que a rede neural de RL apresenta maior complexidade em sua estrutura e funcionamento. Enquanto redes neurais tradicionais focam em mapear entradas diretamente para saídas com base em um conjunto fixo de dados, as redes neurais em RL vão além: elas aprendem por meio da interação contínua entre o agente e o ambiente. Esse processo envolve avaliar ações tomadas em diferentes estados, maximizar recompensas acumuladas ao longo do tempo e ajustar os pesos da rede para otimizar decisões futuras.

A complexidade adicional em redes neurais de RL surge devido à necessidade:

1. **Exploração e Exploração:** O agente deve equilibrar entre explorar novas ações e explorar as melhores ações conhecidas.
2. **Feedback Dinâmico:** A recompensa recebida não é fixa; ela depende das ações anteriores e do estado do ambiente.
3. **Aprendizado Sequencial:** O objetivo é maximizar recompensas a longo prazo, não apenas decisões imediatas.

Portanto, a rede neural baseada em RL exige uma arquitetura que lide com decisões temporais e aprendizado iterativo, o que torna seu funcionamento mais complexo em comparação às redes neurais tradicionais.

4.1 Escalabilidade e Espaços de Estado de Alta Dimensão

Um dos maiores desafios do RL é lidar com espaços de estado e ação extremamente amplos, comuns em jogos complexos. Conforme o tamanho do espaço aumenta, cresce também o custo computacional necessário para explorar e aprender de maneira eficiente. Exemplos como StarCraft II, que envolve milhões de possíveis combinações de estados e ações, evidenciam essa dificuldade (VINYALS et al., 2019).

Para mitigar esse problema, avanços em representações compactas de estado e técnicas como aprendizado por abstração têm sido desenvolvidos. Além disso, o uso de redes neurais profundas para aproximar funções de valor e políticas tem demonstrado resultados promissores, como no caso do Deep Q-Networks (DQN) e variantes como Double DQN e Dueling DQN (MNIH et al., 2015).

4.2 Exploração

Outro dilema central do RL é o equilíbrio entre explorar novas ações para descobrir estratégias melhores e explorar as ações já conhecidas como eficazes. Jogos complexos frequentemente requerem agentes que sejam criativos em suas estratégias para superar adversários humanos ou outros agentes de IA.

Métodos como **curiosity-driven exploration**, em que a recompensa é ajustada para incentivar o agente a explorar estados desconhecidos, têm ganhado destaque. Técnicas baseadas em modelos, que simulam o ambiente para testar estratégias de

forma mais segura e eficiente, também oferecem uma alternativa promissora para resolver esse desafio (SUTTON; BARTO, 2018).

4.3 Robustez em Ambientes Incertos

Muitos jogos e aplicações reais apresentam ambientes estocásticos, nos quais os resultados das ações não são completamente previsíveis. Essa imprevisibilidade pode dificultar o aprendizado e a adaptação de agentes de RL. Um exemplo é o jogo Dota 2, onde as ações de outros jogadores ou agentes afetam diretamente o ambiente e a eficácia das estratégias.

Soluções recentes incluem o uso de **políticas robustas** e treinamento com múltiplos modelos de ambiente, conhecidos como *domain randomization*. Essa abordagem ensina os agentes a lidar com uma ampla gama de variações no ambiente, aumentando sua capacidade de generalização e resiliência (VINYALS et al., 2019).

4.4 Redução de Custos Computacionais

O treinamento de agentes de RL, especialmente quando combinado com aprendizado profundo, é notoriamente intensivo em termos computacionais. O AlphaGo, por exemplo, utilizou milhares de TPUs (unidades de processamento de tensor) durante seu treinamento (SILVER et al., 2016). Isso limita o acesso a tecnologias de ponta para pequenas equipes de pesquisa ou startups.

Métodos mais eficientes, como **treinamento distribuído** e **algoritmos de RL off-policy**, vêm sendo desenvolvidos para reduzir os custos computacionais. Além disso, bibliotecas como TensorFlow e PyTorch têm introduzido ferramentas oti006Dizadas que facilitam a implementação e o treinamento de agentes de RL (GOODFELLOW; BENGIO; COURVILLE, 2016).

4.5 Tendências Emergentes no RL

Os avanços mais recentes apontam para tendências que prometem superar muitas das limitações atuais e expandir ainda mais o potencial do RL:

4.5.1 RL Multiagente

O desenvolvimento de agentes que colaboram ou competem entre si em jogos complexos está se tornando cada vez mais relevante. Jogos como StarCraft II e Dota 2 têm servido como plataformas para estudar interações multiagentes, permitindo o surgimento de estratégias que dependem da comunicação e coordenação em tempo real (VINYALS et al., 2019).

4.5.2 Meta-RL e Transfer Learning

Meta-RL é uma abordagem que permite aos agentes aprenderem a aprender, ou seja, desenvolverem habilidades genéricas que podem ser transferidas para novos ambientes. Isso reduz a necessidade de treinamento intensivo em cada novo problema. Essa técnica é especialmente promissora em jogos que introduzem cenários dinâmicos e variados.

4.5.3 RL com Conscientização Ética

Com o aumento da influência do RL em áreas críticas, questões éticas têm ganhado destaque. Pesquisadores estão investigando maneiras de tornar os agentes de RL mais transparentes em suas decisões, garantindo que os objetivos alinhados com as intenções humanas sejam respeitados.

4.5.4 Integração com Modelos Fundamentais

A integração do RL com modelos generativos, como os modelos de linguagem GPT(Generative Pre-trained Transformer), está emergindo como uma área de pesquisa. Isso abre novas possibilidades, como o uso de RL para guiar diálogos ou tomar decisões baseadas em informações textuais dinâmicas.

4.6 Impactos Futurísticos

O futuro da aprendizagem por reforço em jogos complexos vai além do entretenimento. Tecnologias como o AlphaGo e o AlphaStar demonstraram como as inovações no RL podem influenciar áreas como saúde, onde agentes podem auxiliar

no planejamento de cirurgias, ou na logística, otimizando cadeias de suprimentos globais.

Por meio da pesquisa contínua, espera-se que agentes de RL possam não apenas alcançar, mas superar os limites humanos em criatividade e capacidade de decisão, trazendo inovações que revolucionarão a sociedade como um todo (COSTA, 2019).

A aprendizagem por reforço (RL) demonstrou ser uma tecnologia transformadora, especialmente no domínio dos jogos complexos. Desde os avanços iniciais, como o AlphaGo no Go e o OpenAI Five no Dota 2, até o desenvolvimento de agentes altamente sofisticados como o AlphaStar, ficou evidente que jogos servem como laboratórios valiosos para testar e refinar algoritmos que simulam decisões humanas e exploram estratégias inovadoras.

Através do estudo de RL em jogos, pesquisadores enfrentaram desafios relacionados à escalabilidade, exploração e exploração, robustez em ambientes estocásticos e custos computacionais elevados. Tais obstáculos, embora significativos, impulsionaram inovações tecnológicas importantes, como redes neurais profundas, técnicas de aprendizado por abstração e métodos de treinamento distribuído. Essas soluções não apenas ampliaram a eficácia do RL em ambientes virtuais, mas também possibilitaram sua aplicação em cenários do mundo real.

Mais do que entretenimento, os aprendizados extraídos do RL em jogos estão sendo aplicados em áreas como robótica, saúde, logística e planejamento urbano. Os agentes desenvolvidos em jogos complexos demonstraram competências adaptáveis, abrindo caminhos para melhorias em sistemas de transporte, gerenciamento de recursos e suporte à decisão em ambientes dinâmicos.

Entretanto, desafios ainda persistem. A necessidade de generalização, redução de custos computacionais e maior transparência nos processos de tomada de decisão são questões críticas para o futuro. Além disso, a conscientização ética sobre o impacto social e econômico de agentes autônomos é essencial para que essas tecnologias sejam implementadas de maneira responsável.

Ao longo deste trabalho, ficou evidente que a interseção entre RL e jogos complexos é um campo de estudo estratégico, tanto para a evolução da inteligência artificial quanto para a inovação em múltiplos setores. O contínuo desenvolvimento de técnicas de RL, aliado a uma abordagem ética e acessível, promete um futuro no qual agentes inteligentes poderão resolver problemas de crescente complexidade, beneficiando a sociedade em larga escala.

5 Conclusão

Este trabalho teve como objetivo geral investigar e analisar o uso do Aprendizado por Reforço (RL) como uma abordagem para tornar os jogos mais desafiadores, envolventes e inovadores. A partir de uma revisão bibliográfica abrangente, foi possível entender a evolução do RL e como ele tem sido aplicado com sucesso em jogos complexos, como Go, Dota 2 e StarCraft II, demonstrando seu enorme potencial na criação de agentes de IA que são capazes de aprender e melhorar suas estratégias de forma autônoma.

A avaliação do impacto da aprendizagem por reforço nos jogos complexos revelou que o RL não apenas elevou o nível de desafio para os jogadores humanos, mas também proporcionou novas formas de interação e adaptação, criando ambientes dinâmicos que podem responder a diferentes estilos de jogo. Agentes como o AlphaGo e o OpenAI Five, treinados com grandes volumes de dados, exemplificam como o RL pode transformar a forma como os jogos são jogados, introduzindo uma camada de inteligência que antes não existia.

Ao identificar os principais desafios da aprendizagem por reforço, como a escalabilidade, o custo computacional elevado e a necessidade de maior generalização dos agentes, ficou claro que, embora o RL tenha mostrado avanços significativos, ainda há obstáculos técnicos a serem superados. Soluções para esses desafios, como a utilização de técnicas de meta-RL e RL multiagente, são caminhos promissores para o futuro da tecnologia.

Por outro lado, as vantagens da aprendizagem por reforço para os jogos são evidentes. A principal vantagem é a criação de experiências mais imersivas e personalizadas, com agentes que são capazes de aprender e adaptar-se em tempo real, criando jogos que são não apenas mais desafiadores, mas também mais interessantes e imprevisíveis. A IA aplicada aos jogos se torna, assim, um campo não só de entretenimento, mas de inovação tecnológica, com aplicações que se estendem a outras áreas, como robótica e saúde.

Em resumo, o Aprendizado por Reforço representa um marco significativo na evolução dos jogos e da inteligência artificial. Os avanços discutidos neste trabalho indicam que o RL tem o potencial de tornar os jogos mais desafiadores, envolventes e inovadores, oferecendo oportunidades para a criação de experiências mais ricas e

dinâmicas. Futuros estudos devem se concentrar na superação dos desafios técnicos e na exploração de novas aplicações, tanto dentro quanto fora do universo dos jogos.

Este trabalho contribui para o entendimento de como a aprendizagem por reforço pode ser utilizada para transformar os jogos complexos e abre caminhos para futuras pesquisas que podem expandir suas aplicações em diversas áreas da tecnologia.

5.1 Trabalhos Futuros

Para avançar no uso da aprendizagem por reforço (RL) em jogos complexos, futuros trabalhos podem se concentrar em:

- Aumento da generalização dos agentes treinados em jogos, permitindo sua adaptação a diferentes ambientes e aplicações reais.
- Exploração de abordagens de RL multiagente para melhorar a cooperação e competição entre agentes em jogos e cenários do mundo real.
- Investigação da combinação do RL com outras técnicas de IA, como aprendizado supervisionado, para melhorar a eficiência do treinamento.
- Estudo da ética em RL, com foco em tornar as decisões dos agentes mais transparentes e responsáveis.

Essas áreas podem promover avanços significativos, ampliando o impacto da IA em diversas indústrias e melhorando a eficiência dos sistemas inteligentes.

Referências:

COMI, M. **How to teach AI to play games: Deep Reinforcement Learning**. Disponível em: https://towardsdatascience.com/how-to-teach-an-ai-to-play-games-deep-reinforcement-learning-28f9b920440a. Acesso em: 19 set. 2023.

COSTA, D. B.; LUCCA, G.; ADAMATTI, D. F. **Reinforcement learning and computer games: an approach focused on algorithms analysis**. 2019.

COSTA, Thiago. **Aprendizado por Reforço e suas Aplicações em Jogo**s. Revista Brasileira de Computação Aplicada, v. 11, n. 2, p. 15-27, 2019.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Editora Atlas Ltda., 2017.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning. Cambridge: MIT Press**, 2016.

HASHEMI-POUR, C.; CAREW. **Reinforcement learning**. Disponível em: https://www.techtarget.com/searchenterpriseai/definition/reinforcement-learning#:~:text=Reinforcement%20learning%20is%20a%20machine,learn%20through%20trial%20and%20error.. Acesso em: 19 set. 2023.

MARTINS, C.; PINHEIRO, R.; ZANDONADI, E. **O mercado de jogos eletrônicos e seus impactos na sociedade**. Tese (Sistemas de Informação) – Faculdade Multivix, Cachoeiro de Itapemirim, p. 17, 2018.

MARTINS, Gustavo. **O impacto econômico dos jogos digitais**. São Paulo: PUC-SP, 2018.

MNIH, Volodymyr et al. **Human-level control through deep reinforcement learning**. Nature, v. 518, p. 529-533, 2015.

NEWZOO. **Global Games Market Report 2022**. Disponível em: https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2022-free-version. Acesso em: 13 nov. 2023.

NEWZOO. **Global Games Market Report 2023**. Disponível em: https://newzoo.com. Acesso em: 25 nov. 2024

OPENAI. **OpenAI Five defeats Dota 2 world champions**. Disponível em: https://openai.com/research/openai-five-defeats-dota-2-world-champions. Acesso em: 13 mar. 2023.

OPENAI. **OpenAI Five: AI in Dota 2**. Disponível em: https://openai.com. Acesso em: 25 set. 2024.

SANG-HUN, C**. Google's computer program beats Lee Se-dol in Go tournament**. The New York Times, 15 mar. 2016.

SILVER, David et al. **Mastering the game of Go with deep neural networks and tree search**. Nature, v. 529, p. 484-489, 2016.

SILVER, David et al. **Mastering the game of Go without human knowledge. Nature**, 2017.

SILVER, David**. Reinforcement learning in games**. Nature AI Review, v. 1, n. 1, p. 10-15, 2017.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2. ed. Cambridge: MIT Press, 2018.

VINYALS, Oriol et al. **Grandmaster level in StarCraft II using multi-agent reinforcement learning**. Nature, v. 575, p. 350-354, 2019.

WAZLAWICK, R. S. **Metodologia da pesquisa para ciência da computação**. 2. ed. [S.l.]: Campus, 2014.