PROTOCOLO DE REVISÃO

Pesquisador(a): Oséias Dias de Farias

Tema: Aprendizado por Reforço para Controle Preditivo Adaptativo em Inversores de Frequência: Uma Abordagem sem Modelo para Carregamento Dinâmico de Veículos Elétricos

Objetivos:

- **Levantar e revisar artigos científicos** relevantes sobre a aplicação do Aprendizado por Reforço em sistemas de controle preditivo adaptativo, com ênfase em inversores de frequência e carregamento dinâmico de veículos elétricos.
- **Adquirir conhecimento aprofundado** sobre as principais técnicas de Inteligência Artificial, em particular o Aprendizado por Reforço, e sua aplicação na resolução de problemas similares ao tema proposto, visando construir uma base sólida para a pesquisa.
- **Extrair insights** a partir da análise da literatura, adaptando as abordagens e metodologias encontradas para o desenvolvimento da solução proposta no projeto de pesquisa, focando na inovação e eficiência do controle adaptativo.
- **Identificar lacunas no tema** pesquisado, avaliando a viabilidade de aprofundamento e aplicação do conhecimento adquirido para desenvolvimento de uma tese, contribuindo para a evolução da área e explorando novos horizontes em controle adaptativo e carregamento dinâmico.

Formulação da(s) pergunta(s) da revisão:

- Como usar Aprendizado por Reforço em tempo real para otimizar o processo de controle de inversor de frequência para carregamento eficiente de veículos elétricos?
- Como são desenvolvidos os algoritmos para implementação em bancada desses testes;

Fonte(s):

- IEEE: https://ieeexplore.ieee.org

Data/período da Busca: 01/01/2020 à 27/03/2025

Intervalo de tempo da busca: 5 anos

Palavras-chaves:

Palavra-chave em Português	Sinônimos em Português	Palavra-chave em Inglês	Sinônimos em Inglês
Inversor de frequência		Frequency Inverter	
Veículos elétricos híbridos	carros elétricos	Hybrid electric vehicles	electric cars

Carregamento de carga dinâmica		Dynamic Load Charging	
Controle de corrente		Current control	
Sem Modelo		Model-free	
		Model Predictive Control	MPC
Conversor de fonte de tensão		voltage source converter	
controle eletrônico de potência		power electronics control	
Aprendizagem por reforço		Reinforcement Learning	RL
		Q-Learning	
		Deep Q-Learning	DQN Deep Q-Networks

String(s) de busca utilizada(s):

- **Genérica:**

  ("Model-free" OR "predictive" OR "control" OR "voltage source converter" OR "power electronics control")

  AND

  (RL OR "Reinforcement Learning" OR "MDPs" OR "Q-Learning" OR "DQN" OR "Deep Q-Networks" OR "DQN" OR "Deep Q-Learning")

  AND

  ("EVs" OR "Electric Vehicle*" OR "Frequency Inverter*" OR "Dynamic Load Charging" OR NOT traffic*)

- **IEEE:**

  (((Model-free OR predictive OR control OR voltage source converter OR power electronics control) AND (RL OR Reinforcement Learning OR MDPs OR Q-Learning OR DQN OR Deep Q-Networks OR DQN OR Deep Q-Learning) AND (EVs OR Electric Vehicle* OR Frequency Inverter* OR Dynamic Load Charging) AND NOT traffic))

Critérios de Inclusão e Exclusão

Critérios	Tipo
O artigo deve ter texto completo disponível digitalmente na web.	Inclusão
O artigo deve ter sido publicado nos últimos 5 anos.	Inclusão
O artigo deve ter como ser replicado, detalhando todos os passos necessários para isso.	Inclusão
O artigo deve ter sido escrito em inglês.	Inclusão
O artigo deve ter métodos inovadores de controle aplicado a inversores de frequência com uso de IA.	Inclusão
Artigos publicados em periódicos não confiáveis	Exclusão

Justificativa:

O Controle Preditivo de Modelo (MPC) convencional, embora eficaz, depende de modelos matemáticos precisos do inversor, tornando-o vulnerável a variações dinâmicas e incompatibilidades de parâmetros em sistemas como o carregamento de veículos elétricos (VEs) (J. Rodríguez, 2020). Propõe-se, então, explorar o Aprendizado por Reforço (RL) como alternativa sem modelo para otimizar o controle preditivo em tempo real. Essa abordagem permite que o inversor se adapte autonomamente a condições não lineares e incertas, melhorando eficiência energética e estabilidade sem exigir conhecimento prévio detalhado da planta. O estudo visa contribuir para sistemas de carregamento mais robustos, alinhados à demanda por mobilidade elétrica sustentável.

Critérios de Qualidade

2024	2024
1. Método de controle usando IA inovador.	2
2. Publicado em uma revista de renome.	2
3. Escrito por Doutores de renome na área.	2
4. Possui uma linha de pesquisa alinhada ao objetivo do protocolo de revisão.	2
5. Possui uma metodologia inovadora de estudo de caso.	2

CLASSIFICAÇÃO = Somatória das notas dos critérios (N)
OBS:
N >= 85% (Excelente)
65% =< N <= 85% (Muito Boa)
45% =< N <= 65% (Boa)
25% =< N <= 45% (Média)
N < 25% (Baixa)

Lista dos artigos encontrados

Lista dos artigos incluídos

Nº	Título do artigo	Autores	Publicação	Veículo
1	Model-free Neural Network-based Current Control for Voltage Source Inverter	Oswaldo Menendez Felipe Ruiz Daniel Pesantez Juan Vasconez Jose Rodriguez	2024	2024 IEEE International Conference on Automation/XXVI Congress of the Chilean Association of Automatic Control (ICA-ACCA)
2	Reinforcement Learning-Based Energy Management Control Strategy of Hybrid Electric Vehicles	Fei Chen Peng Mei Hehui Xie Shichun Yang Bin Xu Cong Huang	2022	2022 8th International Conference on Control, Automation and Robotics (ICCAR)
3	Intelligent Voltage Control Method in Active Distribution Networks Based on Averaged Weighted Double Deep Q-network Algorithm	Yangyang Wang Meiqin Mao Liuchen Chang Nikos D. Hatziargyriou	2023	Journal of Modern Power Systems and Clean Energy (Volume: 11, Issue: 1, January 2023)
4	Voltage Regulation in Active Distribution Network with Multiagent Deep Q-Learning Approach	Jianyu Lin	2024	2024 43rd Chinese Control Conference (CCC)
5	Dynamic Tariff Optimization for EV Charging Stations Using Reinforcement Learning	Pooja Jain Ankush Tandon Tushar Soni Yash Soni Vanshika Nirwan Vaidehi Mudgal	2024	2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)
6	Network-Aware Online Charge Control with Reinforcement Learning	Andrey Poddubnyy Phuong Nguyen Han Slootweg	2022	2022 International Conference on Smart Energy Systems and Technologies (SEST)
7	Energy management of hybrid electric vehicles based on model predictive	Chunmei Zhang Wei Cui Yi Du Tao Li Naxin Cui	2022	2022 41st Chinese Control Conference (CCC)

	control and deep reinforcement learning			
8	Reinforcement Learning-based Controller for Thermal Management System of Electric Vehicles	Wansik Choi Jae Woong Kim Changsun Ahn Juhui Gim	2022	2022 IEEE Vehicle Power and Propulsion Conference (VPPC)
9	Assessment of Deep Reinforcement Learning Algorithms for Three-Phase Inverter Control	Oswaldo Menéndez Diana López-Caiza Luca Tarisciotti Felipe Ruiz Fernando Auat-Cheein José Rodríguez	2023	2023 IEEE 8th Southern Power Electronics Conference and 17th Brazilian Power Electronics Conference (SPEC/COBEP)
10	Data-Driven Switching Control Technique Based on Deep Reinforcement Learning for Packed E-Cell as Smart EV Charger	Meysam Gheisarnejad Arman Fathollahi Mohammad Sharifzadeh Eric Laurendeau Kamal Al-Haddad	2024	IEEE Transactions on Transportation Electrification
11	Event-Triggered Model Predictive Control With Deep Reinforcement Learning for Autonomous Driving	Fengying Dang Dong Chen Jun Chen Zhaojian Li	2024	IEEE Transactions on Intelligent Vehicles
12	Distributed Nonlinear Model Predictive Control and Reinforcement Learning	Ifrah Saeed Tansu Alpcan Sarah M. Erfani M. Berkay Yilmaz	2019	2019 Australian & New Zealand Control Conference (ANZCC)
13	Leveraging AI for Enhanced Power Systems Control: An Introductory Study of Model-Free DRL Approaches	Yi Zhou Liangcai Zhou Zhehan Yi Di Shi Mengjie Guo	2024	IEEE
14	A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning	Pouria Qashqai Mohammad Babaie Rawad Zgheib Kamal Al-Haddad	2023	IEEE

15	Reinforcement Learning-Based Predictive Control for Power Electronic Converters	Yihao Wan Qianwen Xu Tomislav Dragičević	2024	IEEE Transactions on Industrial Electronics

N	Título do Artigo	C1	C2	C3	C4	C5	Percentual de qualidade do artigo
	Model-free Neural Network-based Current Control for Voltage Source Inverter	2	2	2	2	2	**100%**
	Reinforcement Learning-Based Energy Management Control Strategy of Hybrid Electric Vehicles	2	2	2	1	1	**80%**
	Intelligent Voltage Control Method in Active Distribution Networks Based on Averaged Weighted Double Deep Q-network Algorithm	2	2	2	2	2	**100%**
	Voltage Regulation in Active Distribution Network with Multiagent Deep Q-Learning Approach	2	2	2	1	1	**80%**
	Dynamic Tariff Optimization for EV Charging Stations Using Reinforcement Learning	2	2	2	2	1	**90%**
	Network-Aware Online Charge Control with Reinforcement Learning	2	2	2	2	2	**100%**
	Energy management of hybrid electric vehicles based on model predictive control and deep reinforcement learning	2	2	2	1	2	**100%**
	Reinforcement Learning-based Controller for Thermal Management System of Electric Vehicles	2	2	2	1	1	**80%**
	Assessment of Deep Reinforcement Learning Algorithms for Three-Phase Inverter Control	2	2	2	2	2	**100%**
	Data-Driven Switching Control Technique Based on Deep Reinforcement Learning for Packed E-Cell as Smart EV Charger	2	2	2	1	2	**90%**

	Event-Triggered Model Predictive Control With Deep Reinforcement Learning for Autonomous Driving	2	2	2	0	1	**70%**
	Distributed Nonlinear Model Predictive Control and Reinforcement Learning	2	2	2	2	2	**100%**
	Leveraging AI for Enhanced Power Systems Control: An Introductory Study of Model-Free DRL Approaches	2	2	2	1	2	**90%**
	A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning	2	2	2	2	2	**100%**
	Reinforcement Learning-Based Predictive Control for Power Electronic Converters	2	2	2	2	2	**100%**

Lista de artigos excluídos

Nº	Título do artigo	Autores	Publicação	Veículo
1				
2				
3				
4				
5				

FORMULÁRIO DE EXTRAÇÃO DE DADOS

1º)

Título do Artigo: Model-free Neural Network-based Current Control for Voltage Source Inverter

Autores: Oswaldo Menendez, Felipe Ruiz, Daniel Pesantez, Juan Vasconez, Jose Rodriguez

Data da Publicação: 09/11/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: This work introduces a current control strategy for Voltage Source Inverters (VSI) using data-driven control systems, particularly employing a framework based on Deep Reinforcement Learning agents. Unlike the other techniques in the literature, we have avoided using a modulator by including a Deep Q-Network agent. In addition, an analysis of the impact of different Deep Neural Network (DNN) architectures on control system performance, specifically considering the number of layers and neurons, is

presented. To this end, different DQN agents were designed, trained, and tested. Also, a two-level voltage source power inverter is simulated to validate the proposed data-driven control based on DQN agents. The performance of the control strategy is analyzed in terms of computational cost, Root Mean Square Error (RMSE), and Total Harmonic Distortion (THD). Simulated results reveal that the proposed control strategy performs strongly in the current control, with a maximum RMSE of 0.83 A and a THD of 5.29% at a 10 kHz sampling frequency when a DNN with one layer and five neurons is used.

Resumo: Este trabalho apresenta uma estratégia de controle de corrente para Inversores de Fonte de Tensão (VSI) usando sistemas de controle orientados por dados, particularmente empregando uma estrutura baseada em agentes de Aprendizado por Reforço Profundo. Diferentemente das outras técnicas na literatura, evitamos usar um modulador incluindo um agente Deep Q-Network. Além disso, é apresentada uma análise do impacto de diferentes arquiteturas de Rede Neural Profunda (DNN) no desempenho do sistema de controle, considerando especificamente o número de camadas e neurônios. Para esse fim, diferentes agentes DQN foram projetados, treinados e testados. Além disso, um inversor de energia de fonte de tensão de dois níveis é simulado para validar o controle orientado por dados proposto com base em agentes DQN. O desempenho da estratégia de controle é analisado em termos de custo computacional, Erro Quadrático Médio (RMSE) e Distorção Harmônica Total (THD). Os resultados simulados revelam que a estratégia de controle proposta tem um forte desempenho no controle atual, com um RMSE máximo de 0,83 A e um THD de 5,29% a uma frequência de amostragem de 10 kHz quando uma DNN com uma camada e cinco neurônios é usada.

Estudo:

Data de execução: 2024
Local: Ambiente Computacional
Tipo: Simulação em MATLAB

Descrição: O artigo propõe uma estratégia de controle de corrente para inversores de fonte de tensão (VSI) utilizando aprendizado por reforço profundo (DRL) com agentes Deep Q-Network (DQN), eliminando a necessidade de moduladores PWM tradicionais. O estudo analisa o impacto de diferentes arquiteturas de redes neurais profundas (DNN) no desempenho do sistema de controle.

Hipóteses avaliadas:

- **Hipótese H0:** Modelos matemáticos são mais eficientes em controle de corrente de inversores de frequências do que qualquer técnicas baseadas em DNN.
- **Hipótese H1:** Um sistema de controle baseado em DQN, sem modelo matemático, pode controlar efetivamente a corrente em um VSI.
- **Hipótese H2**: Arquiteturas mais simples de DNN podem alcançar desempenho comparável ou superior a redes mais complexas, reduzindo custos computacionais.

Variáveis independentes:

- Arquitetura da DNN (número de camadas e neurônios).
- Frequência de amostragem (ex: 10 kHz).
- Parâmetros do agente DQN (taxa de aprendizado, tamanho do buffer de experiência).

Variáveis dependentes:

- RMSE (Erro Quadrático Médio Raiz) da corrente de carga.
- THD (Distorção Harmônica Total) da corrente de carga.
- Tempo de treinamento (custo computacional).
- Tamanho do modelo da DNN (em kB).

Participantes:

Não há participantes humanos. O estudo utiliza simulações computacionais de um VSI de dois níveis conectado a uma carga RL trifásica.

Material:

- Simulador de VSI.
- Agentes DQN com diferentes arquiteturas de DNN (ex: 1 camada com 5 neurônios).
- Métricas de avaliação: RMSE, THD, tempo de treinamento.
- Ambiente de treinamento baseado em framework de DRL (ex: TensorFlow, PyTorch).

Planejamento do estudo:

1. Projeto de agentes DQN com variações na arquitetura da DNN.
2. Treinamento dos agentes em um ambiente simulado de VSI.
3. Avaliação do desempenho por meio de métricas (RMSE, THD) e custo computacional.
4. Comparação entre arquiteturas simples (ex: 1L-5N) e complexas (ex: 2L-50N).

Ameaças à validade:

- Generalização limitada a outros tipos de cargas ou condições operacionais não testadas.
- Dependência da qualidade dos dados de treinamento e da configuração inicial do agente.
- Escalabilidade não verificada para sistemas de maior potência ou com mais fases.

Resultados:

- Arquitetura 1L-5N alcançou RMSE de 0,83 A e THD de 5,29%, com tempo de treinamento de 829 minutos e modelo de 4 kB.

- Arquiteturas mais complexas (ex: 2L-50N) não melhoraram significativamente o desempenho (RMSE: 0,91 A; THD: 5,91%), mas demandaram mais recursos (modelo de 17 kB).
- Redução de 75,16% para 5,13% no THD ao aumentar neurônios em arquiteturas de uma camada.

Comentários adicionais:

- O estudo destaca a viabilidade do DRL em aplicações de eletrônica de potência, com ênfase na otimização de arquiteturas de redes neurais.
- A eliminação do modulador PWM simplifica o sistema, mas a dependência de treinamento intensivo pode limitar aplicações em tempo real.

Referências relevantes

- Rodriguez et al. (2023) sobre controle preditivo sem modelo em VSI.
- Fujimoto et al. (2018) sobre métodos actor-critic para reduzir erros de aproximação.
- Lillicrap et al. (2019) sobre controle contínuo com DRL.

Justificativa:

A pesquisa avança o controle de conversores de potência ao integrar técnicas de inteligência artificial, demonstrando que arquiteturas simples de DNN podem substituir moduladores tradicionais com desempenho satisfatório. Isso reduz a complexidade do sistema e abre caminho para aplicações em microrredes e veículos elétricos, onde eficiência e simplicidade são críticas.

2º)

Título do Artigo: Reinforcement Learning-Based Energy Management Control Strategy of Hybrid Electric Vehicles

Autores: Fei Chen, Peng Mei, Hehui Xie, Shichun Yang, Bin Xu, Cong Huang

Data da Publicação: 31 de Maio de 2022

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: This article is aimed at developing a control strategy based on the Q-learning algorithm for HEVs. The Q-learning algorithm deals with high-dimensional state space problems, and the agent will have a “dimension disaster” problem during the training process. Then a control strategy based on the Deep Q Network (DQN) algorithm is introduced. Since DQN can only output discrete actions, in order to achieve continuous action control, an optimized control strategy based on the Deep Deterministic Policy

Gradient (DDPG) algorithm is proposed. Simulation results show that compared with Q-learning and DQN algorithms, the DDPG algorithm converges faster, and the training process is more robust. Besides, the energy optimization control strategy based on the DDPG algorithm can better control the energy of HEVs.

Resumo: Este artigo tem como objetivo desenvolver uma estratégia de controle baseada no algoritmo Q-learning para HEVs. O algoritmo Q-learning lida com problemas de espaço de estado de alta dimensão, e o agente terá um problema de "desastre dimensional" durante o processo de treinamento. Então, uma estratégia de controle baseada no algoritmo Deep Q Network (DQN) é introduzida. Como o DQN só pode gerar ações discretas, para atingir o controle contínuo de ações, uma estratégia de controle otimizada baseada no algoritmo Deep Deterministic Policy Gradient (DDPG) é proposta. Os resultados da simulação mostram que, em comparação com os algoritmos Q-learning e DQN, o algoritmo DDPG converge mais rápido, e o processo de treinamento é mais robusto. Além disso, a estratégia de controle de otimização de energia baseada no algoritmo DDPG pode controlar melhor a energia dos HEVs.

Estudo

Data de execução: 2022
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O artigo propõe estratégias de controle de gerenciamento de energia para veículos híbridos elétricos (HEVs) utilizando algoritmos de aprendizado por reforço (Q-learning, DQN e DDPG), visando minimizar o consumo de combustível e manter o equilíbrio do estado de carga (SOC) da bateria.

Hipóteses avaliadas:

- O algoritmo DDPG supera Q-learning e DQN em convergência, robustez e eficiência energética.
- Estratégias baseadas em aprendizado por reforço podem otimizar o controle contínuo de ações em HEVs.

Variáveis independentes:

- Algoritmos de aprendizado por reforço (Q-learning, DQN, DDPG).
- Parâmetros de treinamento (taxa de aprendizado α, coeficiente de atenuação γ).
- Condições operacionais do veículo (velocidade v, aceleração a, SOC da bateria).

Variáveis dependentes:

- Consumo acumulado de combustível equivalente.
- Estabilidade do SOC da bateria.
- Tempo de convergência dos algoritmos.

Participantes:

- Modelos computacionais de HEVs (não há participantes humanos).

Material:

- Modelo matemático do sistema híbrido de potência (equações dinâmicas, restrições físicas).
- Redes neurais (DQN: avaliação e alvo; DDPG: Actor-Critic).
- Dados de simulação baseados em ciclos de velocidade reais.
- Ferramentas de normalização de dados e estruturas de priorização (SumTree).

Planejamento do estudo:

- Modelagem do sistema híbrido e definição do problema de otimização.
- Implementação dos algoritmos Q-learning, DQN e DDPG.
- Treinamento offline com ajuste de hiperparâmetros.
- Simulação comparativa usando um ciclo de velocidade pré-definido.
- Análise de convergência, consumo de combustível e desempenho do SOC.

Ameaças à validade:

- Simplificação do modelo do HEV (exclusão de fatores ambientais ou variações não previstas).
- Dependência de dados simulados, que podem não refletir totalmente condições reais.
- Generalização limitada para outros tipos de veículos híbridos ou cenários de condução.

Resultados:

- DDPG: Convergência em ~80 roda1% vs. DQN.
- DQN: Estabilidade após ~500 rodadasdas, redução de 19,71% no consumo de combustível vs. Q-learning e 5,3, menor oscilação que Q-learning.
- Q-learning: Convergência lenta (~4000 rodadas), maior consumo de combustível.

Comentários adicionais:

- O DDPG mostrou-se adequado para controle contínuo em espaços de alta dimensionalidade.
- Sugere-se integração com informações em rede para controle hierárquico futuro.

Referências relevantes

- [9] Qi et al. (2016): Q-learning para otimização de SOC.
- [10] Lin et al. (2015): Aprendizado por reforço aninhado para custos operacionais.
- [14] Hu et al. (2018): Avaliação de estratégias de aprendizado por reforço profundo.

Justificativa:

O DDPG foi escolhido por resolver limitações de Q-learning (tabelas Q) e DQN (ações discretas), permitindo controle contínuo e eficiente em ambientes complexos. Sua arquitetura Actor-Critic e normalização de dados facilitaram a convergência rápida e a otimização do consumo de combustível.

3º)

Título do Artigo: Intelligent Voltage Control Method in Active Distribution Networks Based on Averaged Weighted Double Deep Q-network Algorithm

Autores: Yangyang Wang, Meiqin Mao, Liuchen Chang, Nikos D. Hatziargyriou

Data da Publicação: 27/10/2022

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: High penetration of distributed renewable energy sources and electric vehicles (EVs) makes future active distribution network (ADN) highly variable. These characteristics put great challenges to traditional voltage control methods. Voltage control based on the deep Q-network (DQN) algorithm offers a potential solution to this problem because it possesses human-level control performance. However, the traditional DQN methods may produce overestimation of action reward values, resulting in degradation of obtained solutions. In this paper, an intelligent voltage control method based on averaged weighted double deep Q-network (AWDDQN) algorithm is proposed to overcome the shortcomings of overestimation of action reward values in DQN algorithm and underestimation of action reward values in double deep Q-network (DDQN) algorithm. Using the proposed method, the voltage control objective is incorporated into the designed action reward values and normalized to form a Markov decision process (MDP) model which is solved by the AWDDQN algorithm. The designed AWDDQN-based intelligent voltage control agent is trained offline and used as online intelligent dynamic voltage regulator for the ADN. The proposed voltage control method is validated using the IEEE 33-bus and 123-bus systems containing renewable energy sources and EVs, and compared with the DQN and DDQN algorithms based methods, and traditional mixed-integer nonlinear program based methods. The simulation results show that the proposed method has better convergence and less voltage volatility than the other ones.

Resumo: A alta penetração de fontes de energia renováveis distribuídas e veículos elétricos (VEs) torna a futura rede de distribuição ativa (ADN) altamente variável. Essas características colocam grandes desafios aos métodos tradicionais de controle de tensão. O controle de tensão baseado no algoritmo deep Q-network (DQN) oferece uma solução potencial para esse problema porque possui desempenho de controle de nível humano. No entanto, os métodos tradicionais de DQN podem produzir superestimação dos valores de recompensa de ação, resultando na degradação das soluções obtidas. Neste artigo, um método de controle de tensão inteligente baseado no algoritmo averaged weighted double deep Q-network (AWDDQN) é proposto para superar as deficiências da superestimação dos valores de recompensa de ação no algoritmo DQN e subestimação dos valores de recompensa de ação no algoritmo double deep Q-network (DDQN). Usando o método proposto, o objetivo de controle de tensão é incorporado aos valores de recompensa de

ação projetados e normalizado para formar um modelo de processo de decisão de Markov (MDP) que é resolvido pelo algoritmo AWDDQN. O agente de controle de tensão inteligente baseado em AWDDQN projetado é treinado offline e usado como regulador de tensão dinâmico inteligente online para o ADN. O método de controle de tensão proposto é validado usando os sistemas IEEE 33-bus e 123-bus contendo fontes de energia renováveis e EVs, e comparado com os métodos baseados em algoritmos DQN e DDQN, e métodos tradicionais baseados em programas não lineares inteiros mistos. Os resultados da simulação mostram que o método proposto tem melhor convergência e menos volatilidade de tensão do que os outros.

Estudo

Data de execução: 2023
Local: Ambiente Computacional
Tipo: Simulação no MATLAB

Descrição:

O estudo propõe um método de controle de tensão inteligente em redes de distribuição ativas (ADNs) utilizando o algoritmo Averaged Weighted Double Deep Q-network (AWDDQN). O objetivo é mitigar flutuações de tensão causadas pela alta penetração de fontes renováveis e veículos elétricos (EVs), superando limitações de algoritmos anteriores (DQN e DDQN) e métodos tradicionais baseados em programação não linear inteira mista (MINLP). A eficácia é validada em sistemas IEEE 33-bus e 123-bus com simulações comparativas.

Hipóteses avaliadas:

1. O AWDDQN supera DQN e DDQN na precisão de estimativa de recompensas, evitando superestimação/subestimação.
2. O método proposto é mais eficiente em tempo de cálculo e estabilidade que abordagens tradicionais (MINLP).
3. A capacidade escalonável de EVs (EVSC) pode ser integrada eficazmente no controle de tensão sem comprometer a demanda dos usuários.

Variáveis independentes:

- Algoritmo de controle (AWDDQN, DQN, DDQN, MINLP).
- Parâmetros do sistema (número de barramentos, recursos ajustáveis, perfis de carga, geração renovável).
- Configurações de treinamento (número de episódios, taxa de aprendizado, tamanho da memória).

Variáveis dependentes:

- Volatilidade da tensão (desvio quadrático médio em relação à tensão base).
- Velocidade de convergência do algoritmo.
- Tempo de cálculo para controle online.
- Satisfação da demanda de carga dos EVs (SOC final).

Participantes:

- Sistemas de teste simulados: redes IEEE 33-bus e 123-bus modificadas com integração de fontes renováveis (PV, eólica) e EVs.

Material:

- Ambientes de simulação (MATLAB).
- Algoritmos de aprendizado por reforço profundo (DQN, DDQN, AWDDQN).
- Parâmetros operacionais: eficiência de carregamento/descarga de EVs (η_a=0,98), capacidade da bateria (40 kWh), perfis estocásticos de carga e geração (±10% de flutuação).

Planejamento do estudo:

1. Treinamento offline: Geração de dados de 400 dias via método de Monte Carlo.
2. Teste online: Validação em um dia com flutuações estocásticas.
3. Comparações: AWDDQN vs. DQN, DDQN e MINLP em convergência, desempenho de controle e tempo de execução.
4. Análise de sensibilidade: Efeito do número de ações (K) no treinamento.

Ameaças à validade:

- Validade externa: Resultados limitados aos sistemas IEEE testados; generalização para redes maiores ou mais complexas não foi verificada.
- Validade interna: Simplificação de modelos (ex: comportamento estocástico de EVs e redes) pode subestimar desafios práticos.
- Viés de implementação: Dependência de simulações computacionais, que podem não refletir condições reais de comunicação e latência.

Resultados:

- AWDDQN obteve menor volatilidade de tensão (redução de 71% no IEEE 33-bus e 93% no IEEE 123-bus vs. sem controle).
- Convergência mais estável que DQN/DDQN (episódios: ~1500 para 33-bus, ~7000 para 123-bus).
- Tempo de cálculo online: 0,11% do tempo do MINLP (33-bus) e 0,09% (123-bus).
- EVs: 98% da demanda atendida, com exceção de casos com tempo de conexão curto.

Comentários adicionais:

- O uso de double weighted estimators no AWDDQN mostrou-se eficaz para ambientes estocásticos.
- Sugere-se validar o método em redes com maior variabilidade e integrar aspectos de comunicação em tempo real.

- A escolha de K=8 equilibrou precisão e custo computacional.

Referências relevantes

- Van Hasselt et al. (2016) - Fundamentos do DDQN.
- Zhang et al. (2017) - Modelagem de capacidade escalonável de EVs.
- Wang et al. (2014) - Método MINLP para controle de tensão.

Justificativa:

A integração massiva de fontes renováveis e EVs torna o controle de tensão em ADNs complexo e dinâmico. Métodos tradicionais (ex: MINLP) são lentos e dependentes de modelos, enquanto DQN/DDQN têm viés na estimativa de recompensas. O AWDDQN surge como uma solução robusta e adaptativa, essencial para operação eficiente de redes modernas.

4º)

Título do Artigo: Voltage Regulation in Active Distribution Network with Multiagent Deep Q-Learning Approach

Autores: Jianyu Lin

Data da Publicação: 17/09/2022

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Voltage and reactive power control via inverter-based distributed generators is necessary for distribution power network to mitigate voltage violations. Voltage control is vital to keeping distribution power system voltages within normal range, minimizing power network losses and reducing losses of voltage regulating devices. To deal with incomplete and inaccurate distribution power system models, voltage control problems are solved in a model-free way by using a multi-agent deep reinforcement learning algorithm(DRL). The voltage control problem has been transformed in power system into an deep Q-network(DQN) framework, which avoids solving a specific optimization model directly due to time-varying operating conditions. The designed reward function guides these agents to interact with the distribution power system in the direction of normal voltage interval.

Resumo: O controle de tensão e potência reativa por meio de geradores distribuídos baseados em inversores é necessário para que a rede de distribuição de energia atenue as violações de tensão. O controle de tensão é vital para manter as tensões do sistema de distribuição de energia dentro da faixa normal, minimizando as perdas da rede de energia e reduzindo as perdas dos dispositivos reguladores de tensão. Para lidar com modelos incompletos e imprecisos do sistema de distribuição de energia, os problemas de controle de tensão são resolvidos de forma livre de modelos usando um algoritmo de aprendizado por reforço profundo (DRL) multiagente. O problema de controle de tensão foi

transformado no sistema de energia em uma estrutura de rede Q profunda (DQN), que evita resolver um modelo de otimização específico diretamente devido às condições operacionais variáveis no tempo. A função de recompensa projetada orienta esses agentes a interagir com o sistema de distribuição de energia na direção do intervalo de tensão normal.

Estudo

Data de execução: 2024
Local: Ambiente Computacional
Tipo: Simulação no MATLAB

Descrição:

O artigo propõe um algoritmo de aprendizado por reforço profundo (DRL) multiagente para regulação de tensão em redes de distribuição ativas (ADNs), visando manter as tensões dentro da faixa normal (0,95–1,05 p.u.) sem dependência de modelos precisos do sistema. O método utiliza redes neurais profundas (DNNs) para aproximar a função de valor-Q, permitindo controle adaptativo em condições operacionais variáveis.

Hipóteses avaliadas:

1. Um algoritmo de DRL multiagente pode regular eficazmente a tensão em redes de distribuição sem conhecimento prévio do modelo do sistema.
2. A definição de recompensas baseadas em desvios de tensão e penalizações por violações guia os agentes a otimizar suas políticas de controle.

Variáveis independentes:

- Parâmetros do algoritmo DQN (taxa de aprendizado ν, fator de desconto β
- β, taxa de decaimento d).
- Ações discretas dos agentes (ajustes nas magnitudes de tensão dos geradores, dentro de 0,95–1,05 p.u.).
- Estados do ambiente (medições de tensão, fluxos de potência, cargas).

Variáveis dependentes:

- Estabilidade da tensão nas barras (desvio em relação a Vref=1,0V ref=1,0 p.u.).
- Recompensas acumuladas (positivas para tensões normais, negativas para violações ou divergências).
- Redução de perdas na rede e penalizações por violações de tensão.

Participantes:

- Ambiente simulado do sistema IEEE 200-bus.
- Agentes DRL (controladores de tensão baseados em DQN).

Material:

- Ambiente de simulação em MATLAB com Machine Learning Toolbox.
- Rede neural profunda (DNN) com três camadas totalmente conectadas.

- Conjunto de dados de operação da rede (fluxos de potência, cargas, medições de tensão).

Planejamento do estudo:

1. Treinamento do DQN: Coleta de experiências (estados, ações, recompensas) em um buffer de replay.
2. Atualização da rede neural: Amostragem de mini-lotes para treinar a DNN, minimizando a função de perda.
3. Avaliação de desempenho: Monitoramento das recompensas médias ao longo de 200 episódios de treinamento.
4. Validação: Aplicação do algoritmo no sistema IEEE 200-bus para verificar a regulação de tensão.

Ameaças à validade:

1. Validade externa: Resultados podem não generalizar para redes com topologias ou cargas diferentes.
2. Validade interna: Ambiente simulado simplificado pode não capturar todas as dinâmicas de redes reais.
3. Viés de dados: Dependência da qualidade dos dados de treinamento e parametrização inicial do DQN.

Resultados:

- Durante o treinamento, as recompensas evoluíram de valores negativos (violações frequentes) para positivos (tensões estabilizadas).
- O algoritmo demonstrou capacidade de reduzir violações e manter tensões dentro da faixa normal após iterações suficientes.
- A média de recompensas aumentou continuamente, indicando aprendizado eficaz das políticas de controle.

Comentários adicionais:

- O estudo destaca a vantagem do DQN em espaços de ação discretos, mas sugere futuras extensões para ações contínuas (ex.: uso de soft actor-critic).
- Limitações incluem a necessidade de grandes volumes de dados para treinamento e a complexidade de escalonamento para redes maiores.

Referências relevantes

- Mnih et al. (2015) - Fundamentos do DQN.
- Zhang et al. (2020) - Controle Volt-VAR com DRL.
- Wang et al. (2020) - Algoritmo safe off-policy para controle de tensão.
- Liu e Wu (2021) - Controle multiagente online em redes de distribuição.

Justificativa:

A abordagem é justificada pela complexidade e variabilidade de redes modernas com alta penetração de fontes distribuídas (ex.: solar), onde modelos tradicionais são inviáveis. O DRL oferece uma solução model-free, adaptativa a condições dinâmicas, reduzindo a necessidade de comunicação em tempo real entre dispositivos.

5º)

Título do Artigo: Dynamic Tariff Optimization for EV Charging Stations Using Reinforcement Learning

Autores: Pooja Jain, Ankush Tandon, Tushar Soni, Yash Soni, Vanshika Nirwan, Vaidehi Mudgal

Data da Publicação: 23/10/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: The growing adoption of electric vehicles (EV s) presents both opportunities and challenges for power grids and EV charging stations. This research explores the application of reinforcement learning (RL) for intelligent tariff optimization at charging stations, demonstrating its adaptability to real-world problems. By analyzing real-time data encompassing grid load, hourly electricity prices, and renewable energy availability, our study reveals the RL algorithm's ability to adapt to dynamic conditions and generate tariff schedules that balance grid stability with the seamless integration of renewable energy sources. This work highlights the potential of RL in solving complex energy management challenges and offers valuable insights for energy stakeholders seeking innovative pricing models that align with sustainable energy practices.

Resumo: A crescente adoção de veículos elétricos (VEs) apresenta oportunidades e desafios para redes elétricas e estações de recarga de VEs. Esta pesquisa explora a aplicação do aprendizado por reforço (RL) para otimização de tarifas inteligentes em estações de recarga, demonstrando sua adaptabilidade a problemas do mundo real. Ao analisar dados em tempo real abrangendo carga da rede, preços de eletricidade por hora e disponibilidade de energia renovável, nosso estudo revela a capacidade do algoritmo RL de se adaptar a condições dinâmicas e gerar cronogramas de tarifas que equilibram a estabilidade da rede com a integração perfeita de fontes de energia renováveis. Este trabalho destaca o potencial do RL na solução de desafios complexos de gerenciamento de energia e oferece insights valiosos para as partes interessadas em energia que buscam modelos de precificação inovadores que se alinhem com práticas de energia sustentável.

Estudo

Data de execução: 2024
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O artigo explora a aplicação de aprendizado por reforço (RL) para otimizar tarifas dinâmicas em estações de carregamento de veículos elétricos (EV). O estudo utiliza dados em tempo real, como carga da rede elétrica, preços horários de eletricidade e disponibilidade de energia renovável, para ajustar as tarifas de forma inteligente, equilibrando a estabilidade da rede com a integração de fontes de energia renovável.

Hipóteses avaliadas:

O uso de algoritmos de aprendizado por reforço pode melhorar a programação dinâmica de tarifas, ajustando as tarifas com base nas condições em tempo real da rede e promovendo o uso eficiente da energia renovável.

Variáveis independentes:

- Carga da rede elétrica
- Preços horários de eletricidade
- Disponibilidade de energia solar

Variáveis dependentes:

- Tarifas de carregamento dos veículos elétricos
- Comportamento de carga dos veículos (deslocamento de carga)
- Estabilidade da rede elétrica

Participantes:

O estudo é baseado em dados simulados e não envolve participantes humanos diretamente, mas utiliza perfis de demanda de carregamento de EV, geração solar e carga da rede.

Material:

- Dados de carga horária da rede elétrica
- Dados de geração de energia solar de uma planta fotovoltaica de 400 kW
- Perfis de demanda de carregamento de EV

Planejamento do estudo:

Foi utilizado um modelo de aprendizado por reforço (Proximal Policy Optimization - PPO) para otimizar as tarifas dinâmicas. A pesquisa envolveu a simulação de diferentes cenários de carga da rede, disponibilidade solar e demanda de carregamento de EV para ajustar as tarifas em tempo real.

Ameaças à validade:

- Limitações de dados simulados e não reais
- Escalabilidade dos modelos de aprendizado por reforço em implementações reais
- Desafios na integração de dados em tempo real

Resultados:

O modelo de aprendizado por reforço foi capaz de ajustar as tarifas de maneira dinâmica, reduzindo a demanda de pico de carregamento em até 20% dependendo da elasticidade do preço, o que ajudou a mitigar a sobrecarga da rede elétrica.

Comentários adicionais:

O estudo mostra o potencial do aprendizado por reforço para otimizar o gerenciamento de tarifas em estações de carregamento de EV, mas reconhece que mais pesquisas são necessárias para validar esses resultados em cenários do mundo real.

Referências relevantes:

Justificativa: A pesquisa é relevante pois aborda um problema crescente de como integrar eficientemente os veículos elétricos à rede elétrica, otimizando o uso de energia renovável e mantendo a estabilidade da rede. A utilização de algoritmos de aprendizado por reforço oferece uma solução adaptável e eficiente para a programação dinâmica das tarifas.

6º)

Título do Artigo: Network-Aware Online Charge Control with Reinforcement Learning

Autores: Andrey Poddubnyy, Phuong Nguyen, Han Slootweg

Data da Publicação: 28/09/2022

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: he rising market of electric vehicles leads to an increased number of chargers needed to be connected to the power grid. It sometimes leads to their installation being postponed due to the capacity limits violations, which can be eased by implementing a smart charging control. In this paper the smart charging control is proposed from the point of view of grid loading as a primary indicator. The reinforcement learning agent controls the charger's power of consumption to optimize expenses and prevent lines and transformers from being overloaded. The simulations were carried out in the IEEE 13 bus test feeder with the load profile data coming from the residential area. To simulate the real availability of data, an agent is trained with only the transformer current and the local charger's state, like state of the charge and timestamp. Several algorithms are tested to select the best one to utilize in the stochastic environment and low frequency of data streaming.

Resumo: O mercado crescente de veículos elétricos leva a um aumento no número de carregadores necessários para serem conectados à rede elétrica. Às vezes, isso leva ao adiamento de sua instalação devido às violações dos limites de capacidade, o que pode ser facilitado pela implementação de um controle de carregamento inteligente. Neste artigo, o controle de carregamento inteligente é proposto do ponto de vista do carregamento da rede como um indicador primário. O agente de aprendizado por reforço controla a potência de consumo do carregador para otimizar as despesas e evitar que linhas e transformadores sejam sobrecarregados. As simulações foram realizadas no alimentador de teste de barramento IEEE 13 com os dados do perfil de carga provenientes da área residencial. Para simular a disponibilidade real dos dados, um agente é treinado apenas com a corrente do transformador e o estado do carregador local, como estado da carga e registro de data e hora. Vários algoritmos são testados para selecionar o melhor para utilizar no ambiente estocástico e baixa frequência de streaming de dados.

Estudo

Data de execução: 2022
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O estudo propõe um algoritmo de controle de carga inteligente para veículos elétricos (VEs) utilizando aprendizado por reforço (RL), com foco na otimização de custos e prevenção de sobrecargas em redes de distribuição. O algoritmo considera o estado da rede (correntes em transformadores/linhas) e dados locais do carregador (estado de carga, tempo).

Hipóteses avaliadas:

1. Algoritmos de RL podem otimizar custos de carregamento e evitar sobrecargas em redes elétricas.
2. Algoritmos model-based (ex: Dyna-Q) têm desempenho semelhante a model-free (ex: Q-learning) em ambientes estocásticos com baixa frequência de dados.

Variáveis independentes:

- Tipo de algoritmo de RL (SARSA, Q-learning, Dyna-Q).
- Parâmetros de RL (taxa de aprendizado α, fator de desconto γ).
- Dados de carga residencial (perfis diários, picos sazonais).
- Configuração da rede elétrica (IEEE 13 bus).

Variáveis dependentes:

- Recompensa cumulativa por episódio.
- Custos de energia.
- Carga percentual em transformadores/linhas.
- Estado de carga (SOC) da bateria do VE.

Participantes:

- Dados de consumo residencial de dois anos (2016–2017), fornecidos por parceiros industriais.
- Rede elétrica simulada (IEEE 13 bus test feeder).

Material:

- Ambiente de simulação em Python.
- Biblioteca PandaPower para simulação de fluxo de carga.
- Bateria de VE simulada (100 kWh, 40 kW de potência).
- Dados de preços de energia e penalidades por sobrecarga.

Planejamento do estudo:

1. Modelagem do problema como um Processo de Decisão Markoviano (MDP).
2. Implementação de três algoritmos de RL (SARSA, Q-learning, Dyna-Q).
3. Simulação em rede IEEE 13 bus com dados reais de carga residencial.
4. Análise comparativa de métricas (recompensa cumulativa, componentes do custo, SOC).
5. Testes de sensibilidade para α e γ.

Ameaças à validade:

- Simplificação excessiva das ações (apenas "carregar" ou "não carregar").
- Suposição de disponibilidade contínua do VE para carregamento.
- Limitação da rede teste (IEEE 13 bus), podendo não representar redes complexas.
- Dependência de dados históricos, sem variações imprevistas em tempo real.

Resultados:

- Todos os algoritmos convergiram para desempenho similar, com Q-learning mostrando adaptação mais estável.
- O agente aprendeu a priorizar períodos de baixo custo e reduzir sobrecargas, mas sacrificou o SOC em momentos de alto custo/sobrecarga.
- Valores altos de γ (0.9) e α moderado (0.1) otimizaram a recompensa.
- Dyna-Q não superou algoritmos model-free, indicando suficiência de dados para aprendizado direto.

Comentários adicionais:

- O estudo demonstra que algoritmos simples de RL são viáveis para controle de carga, reduzindo a necessidade de redes neurais complexas.
- Sugere-se explorar cenários multiagente e redes maiores em trabalhos futuros.

Referências relevantes

- Sutton & Barto (2018) - Fundamentos de RL.
- Mocanu et al. (2019) - RL para otimização energética em edifícios.
- Abdullah et al. (2021) - Revisão de RL aplicado a VEs.

Justificativa:

O estudo é relevante para acelerar a implantação de infraestrutura de VEs sem custos elevados de reforço de rede. A integração de RL com dados de rede elétrica oferece uma abordagem inovadora para gestão de congestão, alinhando-se a demandas de sustentabilidade e eficiência energética.

7º)

Título do Artigo: Energy management of hybrid electric vehicles based on model predictive control and deep

Autores: Chunmei Zhang, Wei Cui, Yi Du, Tao Li, Naxin Cui

Data da Publicação: 11/10/2022

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: In this paper, a model predictive control-deep reinforcement learning (MPC-DRL) based energy management strategy (EMS) is proposed for hybrid electric vehicles (HEVs) to realize optimal energy distribution. Firstly, the Bi-directional Long Short-Term Memory (Bi-LSTM) network is used to predict the vehicle velocity sequence over the prediction horizon. Then, a battery state-of-charge (SOC) reference trajectory planning model is constructed based on vehicle velocity changes and driving mileage. Finally, according to the predicted vehicle velocity and the SOC reference trajectory, the deep Q network (DQN) algorithm searches for optimal control solutions under the MPC framework. The simulation results show that the Bi-LSTM network can accurately predict the vehicle velocity and the MPC-DRL strategy can better follow the downward trend of the SOC reference trajectory. The proposed strategy can achieve similar fuel economy performance to the MPC-dynamic programming (DP) strategy and approximately 9.58% promotion than the ECMS strategy.

Resumo: Neste artigo, uma estratégia de gerenciamento de energia (EMS) baseada em controle preditivo de modelo-aprendizagem por reforço profundo (MPC-DRL) é proposta para veículos elétricos híbridos (HEVs) para realizar a distribuição ideal de energia. Primeiramente, a rede Bi-directional Long Short-Term Memory (Bi-LSTM) é usada para prever a sequência de velocidade do veículo ao longo do horizonte de previsão. Então, um modelo de planejamento de trajetória de referência do estado de carga da bateria (SOC) é construído com base nas mudanças de velocidade do veículo e na quilometragem de direção. Finalmente, de acordo com a velocidade prevista do veículo e a trajetória de referência do SOC, o algoritmo deep Q network (DQN) busca soluções de controle ideais sob a estrutura do MPC. Os resultados da simulação mostram que a rede Bi-LSTM pode prever com precisão a velocidade do veículo e a estratégia MPC-DRL pode seguir melhor a tendência de queda da trajetória de referência do SOC. A estratégia proposta pode atingir desempenho de economia de combustível semelhante à estratégia de programação dinâmica (DP) do MPC e aproximadamente 9,58% de promoção do que a estratégia ECMS.

Estudo

Data de execução: 2022
Local: Ambiente Computacional
Tipo: Simulação no MATLAB

Descrição:

O artigo propõe uma estratégia de gestão de energia para veículos híbridos elétricos (HEVs) que integra Model Predictive Control (MPC) e Deep Reinforcement Learning (DRL). Utiliza uma rede Bi-LSTM para prever a velocidade do veículo, planeja uma trajetória de referência para o estado de carga (SOC) da bateria e aplica o algoritmo Deep Q Network (DQN) no framework MPC para otimizar a distribuição de torque entre o

motor a combustão e o motor elétrico, visando economia de combustível e estabilidade do SOC.

Hipóteses avaliadas:

1. A combinação de MPC com DRL (especificamente DQN) melhora a economia de combustível em comparação com estratégias como ECMS.
2. A previsão de velocidade com Bi-LSTM é mais precisa do que métodos tradicionais (ex: LSTM).
3. O planejamento adaptativo da trajetória de SOC baseado na velocidade prevista permite um controle mais eficiente.

Variáveis independentes:

- Velocidade prevista do veículo (via Bi-LSTM).
- Trajetória de referência do SOC.
- Parâmetros do modelo do veículo (massa, coeficiente de resistência do ar, etc.).
- Hiperparâmetros das redes neurais (Bi-LSTM e DQN).

Variáveis dependentes:

- Consumo de combustível (L/100 km).
- Precisão da previsão de velocidade (MAE, RMSE).
- Capacidade de seguir a trajetória de SOC.
- Tempo de computação do algoritmo.

Participantes:

- Modelo de simulação de um HEV de eixo único paralelo (dados técnicos descritos no artigo).

Material:

- Modelo dinâmico do HEV (equações de movimento, motor, motor elétrico, bateria).
- Rede Bi-LSTM para previsão de velocidade.
- Algoritmo DQN integrado ao MPC.
- Dados de ciclos de condução para treinamento e teste.

Planejamento do estudo:

1. Previsão de velocidade: Treinamento da Bi-LSTM com dados históricos de velocidade.
2. Planejamento do SOC: Definição da trajetória de referência baseada na velocidade prevista e na quilometragem total.
3. Otimização MPC-DRL: Implementação do DQN para distribuição de torque, com restrições do SOC.
4. Comparação: Análise de desempenho contra ECMS e MPC-DP em termos de economia de combustível e seguimento do SOC.

Ameaças à validade:

- Dependência da qualidade dos dados de treinamento para a Bi-LSTM.

- Simplificação do modelo do veículo (ex: desconsideração de variações ambientais ou degradação da bateria).
- Generalização limitada para ciclos de condução não testados ou condições de tráfego dinâmicas.

Resultados:

- Bi-LSTM vs LSTM: Menores erros de previsão (MAE: 0,1433 vs 0,2401; RMSE: 0,7533 vs 0,9683).
- Economia de combustível: MPC-DRL reduziu o consumo em 9,58% comparado ao ECMS e teve desempenho próximo ao MPC-DP.
- SOC: MPC-DRL seguiu a trajetória de referência, mas com desempenho ligeiramente inferior ao MPC-DP.

Comentários adicionais:

A abordagem combina técnicas avançadas de controle e aprendizado de máquina, demonstrando potencial para aplicações práticas. Entretanto, desafios como implementação em tempo real e adaptação a cenários complexos (ex: tráfego urbano dinâmico) precisam ser explorados.

Referências relevantes

- ECMS, MPC, DP (referências [6], [7], [8]).
- Aprendizado por reforço aplicado a HEVs ([14], [15]).
- Trabalhos anteriores em previsão de velocidade ([9], [10], [11]).

Justificativa:

A estratégia proposta busca superar limitações de métodos tradicionais (ex: regras fixas ou otimização global com alto custo computacional) ao integrar a previsão adaptativa (Bi-LSTM) e a otimização baseada em aprendizado (DQN). Isso permite melhor equilíbrio entre eficiência energética, realismo computacional e adaptação a condições variáveis.

8º)

Título do Artigo: Reinforcement Learning-based Controller for Thermal Management System of Electric Vehicles

Autores: Wansik Choi, Jae Woong Kim, Changsun Ahn, Juhui Gim

Data da Publicação: 05/01/2023

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: O sistema de gerenciamento térmico em veículos elétricos (VE) se torna significativo porque o desempenho do sistema é altamente correlacionado com a autonomia, confiabilidade e segurança dos veículos elétricos. Portanto, um controlador do sistema térmico deve ser projetado para minimizar o erro de rastreamento e o consumo de energia, ao mesmo tempo em que satisfaz as restrições. Neste estudo, um controlador baseado em aprendizado por reforço (RL) é proposto. Este artigo apresenta a seleção de estados e o design da função de recompensa de RL para o sistema de gerenciamento térmico de VE. O controlador é treinado pelo método de aprendizado sequencial baseado em Deep Q-network (DQN) e ajustado para convergência rápida de problemas de múltiplas entradas. Os resultados mostram melhor desempenho em comparação com o controlador baseado em regras.

Resumo: O sistema de gerenciamento térmico em veículos elétricos (VE) se torna significativo porque o desempenho do sistema é altamente correlacionado com a autonomia, confiabilidade e segurança dos veículos elétricos. Portanto, um controlador do sistema térmico deve ser projetado para minimizar o erro de rastreamento e o consumo de energia, ao mesmo tempo em que satisfaz as restrições. Neste estudo, um controlador baseado em aprendizado por reforço (RL) é proposto. Este artigo apresenta a seleção de estados e o design da função de recompensa de RL para o sistema de gerenciamento térmico de VE. O controlador é treinado pelo método de aprendizado sequencial baseado em Deep Q-network (DQN) e ajustado para convergência rápida de problemas de múltiplas entradas. Os resultados mostram melhor desempenho em comparação com o controlador baseado em regras.

Estudo

Data de execução: 2022
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

Estudo propõe um controlador baseado em Reinforcement Learning (RL) para o sistema de gerenciamento térmico de veículos elétricos (EVs). Utiliza Deep Q-Network (DQN) com aprendizado sequencial para otimizar o controle da velocidade do compressor e do ventilador de resfriamento, visando minimizar o erro de temperatura da cabine e o consumo de energia, respeitando restrições operacionais. Resultados demonstram superioridade frente a controladores baseados em regras.

Hipóteses avaliadas:

- Controladores RL podem otimizar simultaneamente o desempenho térmico e a eficiência energética em EVs.
- O método DQN com aprendizado sequencial converge mais rapidamente em ambientes de múltiplas entradas.
- A função de recompensa projetada garante o cumprimento implícito das restrições do sistema.

Variáveis independentes:

- Velocidade do compressor (RPM/s).
- Velocidade do ventilador de resfriamento (RPM/s).
- Condições ambientais (temperatura, umidade, intensidade solar).
- Ciclos de condução (US06, SC03).

Variáveis dependentes:

- Erro absoluto da temperatura da cabine.
- Consumo de energia do compressor, ventilador e aquecedor PTC.
- Temperatura do ar no evaporador.
- Pressão do condensador.

Participantes:

- Modelo computacional de um EV médio (Simulink), com bateria de 64 kWh, volume de cabine de 2,66 m³ e peso de 1.685 kg.

Material:

- Modelo térmico do EV desenvolvido em Simulink.
- Framework de RL em Python com TensorFlow para treinamento das redes neurais.
- Comunicação TCP/IP entre Simulink (modelo) e Python (controle).

Planejamento do estudo:

1. **Treinamento**:
    - Condições ambientais e de condução variáveis (distribuição uniforme).
    - Aplicação de ciclos de condução US06 e SC03.
    - Treinamento sequencial de duas redes DQN (compressor e ventilador).
1. **Testes:**
    - Validação em três cenários fixos (temperaturas ambiente de 24°C, 32°C e 40°C).
    - Comparação com controlador baseado em regras.

Ameaças à validade:

- Validade interna: Dependência da precisão do modelo Simulink.
- Validade externa: Generalização limitada a outros cenários não testados (ex.: climas extremos).
- Viés de implementação: Uso de aproximações na função de recompensa (ex.: consumo do ventilador estimado).

Resultados:

- Redução do erro absoluto médio da temperatura da cabine.
- Consumo de energia 15-20% menor comparado ao controlador baseado em regras.

- Resposta rápida do compressor para reduzir erros iniciais, seguida de ajustes para economia energética.

Comentários adicionais:

- Método inovador ao aplicar DQN sequencial para múltiplas entradas, mas carece de validação em sistemas físicos.
- Função de recompensa multifacetada (5 termos) é crucial para equilibrar desempenho e restrições.

Referências relevantes

- Mnih et al. (2015): Base teórica do DQN.
- Wang et al. (2016): Arquitetura "Dueling Network" para RL.
- Amini et al. (2019, 2020): Controle preditivo aplicado a sistemas térmicos.

Justificativa:

O RL é adequado para sistemas complexos como o gerenciamento térmico de EVs, onde modelos físicos precisos são difíceis de obter. A abordagem permite otimização multi-objetivo (desempenho térmico, eficiência energética) através de uma função de recompensa bem projetada, incorporando restrições de forma implícita. O aprendizado sequencial acelera a convergência em ambientes de múltiplas ações, tornando a solução viável para aplicações práticas.

9º)

Título do Artigo: Assessment of Deep Reinforcement Learning Algorithms for Three-Phase Inverter Control

Autores: Oswaldo Menéndez, Diana López-Caiza, Luca Tarisciotti, Felipe Ruiz, Fernando Auat-Cheein, José Rodríguez

Data da Publicação: 01/02/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Deep reinforcement learning (DRL) offers outstanding algorithms to develop optimal controllers for power converters with uncertainties and non-linear dynamics. This work comprehensively analyses a model-free control algorithm for three-phase inverters using DRL agents. To this end, different deep deterministic policy gradient (DDPG) agents with variable hyperparameters were conceptualized, designed, and tested. On average, DDPG agents were shown to have excellent performance in the control of power inverters. Indeed, DDPG agents reduce the impact of model uncertainties and non-linear

dynamics. To validate the proposed control policy, the two-level voltage source power inverter is simulated. Also, the main features of the control strategy are analyzed in terms of computational cost, root medium square error (RMSE), and total harmonic distortion (THD). Simulated results reveal that the proposed control strategy exhibits strong performance in the current control task, achieving a maximum RMSE of 0.78 A and a THD of 3.17% for a 6 kHz sampling frequency.

Resumo: O aprendizado por reforço profundo (DRL) oferece algoritmos excelentes para desenvolver controladores ideais para conversores de energia com incertezas e dinâmica não linear. Este trabalho analisa de forma abrangente um algoritmo de controle sem modelo para inversores trifásicos usando agentes DRL. Para esse fim, diferentes agentes de gradiente de política determinística profunda (DDPG) com hiperparâmetros variáveis foram conceituados, projetados e testados. Em média, os agentes DDPG demonstraram ter excelente desempenho no controle de inversores de energia. De fato, os agentes DDPG reduzem o impacto das incertezas do modelo e da dinâmica não linear. Para validar a política de controle proposta, o inversor de energia de fonte de tensão de dois níveis é simulado. Além disso, as principais características da estratégia de controle são analisadas em termos de custo computacional, erro quadrático médio (RMSE) e distorção harmônica total (THD). Os resultados simulados revelam que a estratégia de controle proposta exibe forte desempenho na tarefa de controle de corrente, atingindo um RMSE máximo de 0,78 A e um THD de 3,17% para uma frequência de amostragem de 6 kHz.

Estudo

Data de execução: 2023
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

Análise de algoritmos de Deep Reinforcement Learning (DRL), especificamente o Deep Deterministic Policy Gradient (DDPG), para controle de inversores trifásicos de tensão com carga RL. O estudo avalia desempenho, custo computacional e adaptabilidade do controle proposto.

Hipóteses avaliadas:

- Agentes DDPG são eficazes para controle de inversores trifásicos, reduzindo o impacto de incertezas do modelo e dinâmicas não lineares.
- A estratégia de controle sem modelo baseada em DRL pode alcançar baixo RMSE e THD, mesmo sob variações de carga e referência de corrente.

Variáveis independentes:

- Tempo de amostragem (100 μs a 500 μs).
- Frequência de comutação (2 kHz a 10 kHz).
- Função de recompensa (pesos k1, k2, k3).
- Hiperparâmetros de treinamento (tamanho do buffer de experiência, taxa de desconto, arquitetura das redes neurais).

Variáveis dependentes:
- RMSE (Erro Quadrático Médio Raiz) da corrente de fase.
- THD (Distorção Harmônica Total) da corrente.
- Tempo de treinamento.
- Custo computacional.

Participantes:
- Ambiente simulado de um inversor trifásico com carga RL.
- Agentes DRL (DDPG) treinados no MATLAB.

Material:
- Software: MATLAB 2023a com Reinforcement Learning Toolbox.
- Hardware: Computador com processador Intel i7-10700F, GPU NVIDIA RTX 4070 e 64GB de RAM.
- Modelo do inversor trifásico (VSI) descrito no artigo.

Planejamento do estudo:

1. Projeto de agentes DDPG com variação de hiperparâmetros.
2. Treinamento dos agentes em diferentes cenários (variação de carga, frequência de comutação e referência de corrente).
3. Avaliação de métricas (RMSE, THD, tempo de treinamento) em estado estacionário e dinâmico.
4. Comparação com estratégias de controle tradicionais (ex: FCS-MPC).

Ameaças à validade:

- Generalização limitada para outros tipos de inversores ou cargas não testadas.
- Dependência de simulações computacionais (não validação experimental).
- Escolha heurística de hiperparâmetros pode introduzir viés.

Resultados:

- Melhor desempenho: RMSE de 0,78 A e THD de 3,17% com 6 kHz de frequência de amostragem.
- Robustez: Adaptação a variações de ±10% na carga e mudanças na referência de corrente.
- Custo computacional: Tempo de treinamento de até 2333 minutos (20 kHz).

Comentários adicionais:

- Contribuição relevante para controle de inversores sem modelo, mas alto custo computacional limita aplicação em tempo real.
- Sugere-se explorar funções de recompensa mais eficientes e otimização de hiperparâmetros em trabalhos futuros.

Referências relevantes

- [4] Zhao et al. (2021): Visão geral de IA em eletrônica de potência.
- [37] Liu et al. (2023): Combinação de FCS-MPC com DRL.
- [42] Xiang et al. (2022): Controle de inversores trifásicos usando DDPG.
- [59] Lillicrap et al. (2019): Fundamentos do algoritmo DDPG.

Justificativa:

O estudo justifica-se pela necessidade de controladores adaptáveis para sistemas de energia com dinâmicas complexas, onde métodos tradicionais (ex: controle linear) têm limitações em eficiência e flexibilidade. A abordagem com DRL visa superar essas barreiras, oferecendo autocalibração e tolerância a incertezas.

10º)

Título do Artigo: Data-Driven Switching Control Technique Based on Deep Reinforcement Learning for Packed E-Cell as Smart EV Charger

Autores: Meysam Gheisarnejad, Arman Fathollahi, Mohammad Sharifzadeh, Eric Laurendeau, Kamal Al-Haddad

Data da Publicação: 30/07/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Among hybrid multilevel rectifiers (HMRs), packed E-cell appeared as an interesting topology due to the generation of nine-level voltage with minimum active/passive devices, but appropriate control design of PEC rectifier is vital demand to keep capacitors voltages well-regulated even under unbalanced/variable dc loads. Therefore, the backstepping control (BSC) strategy is developed to control a nine-level packed E-cell (PEC9) rectifier to be used as a smart EV charger. Proximal policy optimization (PPO) with actor and critic deep neural networks (ADNNs and CDNNs) is trained to adjust the BSC controller, where the PEC9 rectifier can intelligently deal with asymmetrical/symmetrical dc loads. By maximizing a reward function, the PPO agent tries to find the optimal policy to design the control coefficients of BSC with the aim of regulating the PEC9 capacitor's voltages. The developed BSC based on the PPO tuner is

validated using hardware-in-the-loop (HiL) and experimental implementation of the PEC9 rectifier to assess the performance of the proposed control scheme.

Resumo: Entre os retificadores híbridos multinível (HMRs), a célula E empacotada apareceu como uma topologia interessante devido à geração de tensão de nove níveis com dispositivos ativos/passivos mínimos, mas o projeto de controle apropriado do retificador PEC é uma demanda vital para manter as tensões dos capacitores bem reguladas, mesmo sob cargas CC desbalanceadas/variáveis. Portanto, a estratégia de controle de retrocesso (BSC) é desenvolvida para controlar um retificador de célula E empacotada de nove níveis (PEC9) para ser usado como um carregador EV inteligente. A otimização de política proximal (PPO) com redes neurais profundas de ator e crítico (ADNNs e CDNNs) é treinada para ajustar o controlador BSC, onde o retificador PEC9 pode lidar de forma inteligente com cargas CC assimétricas/simétricas. Ao maximizar uma função de recompensa, o agente PPO tenta encontrar a política ótima para projetar os coeficientes de controle do BSC com o objetivo de regular as tensões do capacitor PEC9. O BSC desenvolvido com base no sintonizador PPO é validado usando hardware-in-the-loop (HiL) e implementação experimental do retificador PEC9 para avaliar o desempenho do esquema de controle proposto.

Estudo

Data de execução: 2024
Local: Ambiente Computacional e Bancada de Laboratório
Tipo: Simulação no MATLAB e Teste em Bancada

Descrição:

O artigo propõe uma técnica de controle de comutação baseada em aprendizado por reforço profundo (DRL) para um retificador multinível Packed E-Cell de nove níveis (PEC9), visando sua aplicação como carregador inteligente para veículos elétricos (EVs). O método combina controle backstepping (BSC) com o algoritmo de otimização proximal de política (PPO) para regular as tensões dos capacitores do retificador, mesmo sob cargas CC desbalanceadas ou variáveis.

Hipóteses avaliadas:

1. O ajuste dos coeficientes do controlador BSC via PPO melhora a regulação das tensões dos capacitores do PEC9.
2. A abordagem model-free do BSC, combinada com PPO, é robusta contra dinâmicas não modeladas, incertezas e variações de carga.
3. O PEC9 com controle proposto pode operar eficientemente como carregador de EVs, mantendo qualidade de energia (baixa THD) e correção de fator de potência.

Variáveis independentes:

- Parâmetros do controlador BSC (ω, γ, ψ) ajustados pelo PPO.
- Condições operacionais (cargas CC simétricas/assimétricas, variações na tensão da rede).
- Frequência de comutação e tempo de amostragem (20 μs).

Variáveis dependentes:

- Tensões dos capacitores Vc1, Vc2, Vc3.
- Distorção harmônica total (THD) da tensão de saída.
- Corrente e tensão da rede (sincronização e fator de potência).

Participantes:

Não se aplica (estudo técnico com simulações hardware-in-the-loop (HIL) e protótipo experimental).

Material:

- Retificador PEC9 com três capacitores, seis chaves normais e uma bidirecional.
- Controlador BSC ajustado por PPO (com redes neurais actor e critic).
- Plataforma OPAL-RT para testes HIL.
- Protótipo experimental com cargas resistivas e fontes CC.

Planejamento do estudo:

1. Simulações HIL:
    - Cenário I: Variações simétricas e assimétricas nas cargas CC.
    - Cenário II: Variações na tensão da rede (queda de 10%).
2. Testes experimentais:
    - Validação do protótipo sob carga resistiva fixa (80 Ω).
3. Comparação com controle BSC convencional (sem PPO).

Ameaças à validade:

- Generalização limitada a outros topologias de retificadores multinível.
- Condições operacionais não testadas (e.g., cargas não lineares ou transitórios extremos).
- Dependência da precisão do modelo de rede neural no PPO.

Resultados:

- Cenário I: Tensões dos capacitores mantidas em
- Vc1=180V, Vc2=Vc3=45 V, com THD de 3,2% (vs. 4,3% no BSC convencional).
- Cenário II: Regulação estável das tensões durante queda de tensão na rede.
- Validação experimental: Tensões balanceadas e forma de onda de nove níveis estável, com sincronização entre tensão e corrente da rede.

Comentários adicionais:

- O uso de PPO para ajustar parâmetros do BSC demonstra sinergia entre métodos clássicos e aprendizado de máquina.
- A abordagem model-free reduz a dependência de identificação precisa do sistema, favorecendo aplicações práticas.

Referências relevantes

- Liu et al. (2021) sobre retificadores modulares híbridos.
- Zhang et al. (2021) sobre controle backstepping em retificadores.
- Schulman et al. (2017) sobre o algoritmo PPO.

Justificativa:

O artigo justifica a abordagem pela necessidade de controladores robustos e adaptativos para retificadores multinível em aplicações de carregamento de EVs, onde cargas desbalanceadas e variações dinâmicas são comuns. A combinação de BSC e PPO supera limitações de métodos convencionais (e.g., MPC e SMC), oferecendo regulação precisa sem dependência de modelos matemáticos complexos.

11º)

Título do Artigo: Event-Triggered Model Predictive Control With Deep Reinforcement Learning for Autonomous Driving

Autores: Fengying Dang, Dong Chen, Jun Chen, Zhaojian Li

Data da Publicação: 03/11/2023

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Event-triggered model predictive control (eMPC) is a popular optimal control method with an aim to alleviate the computation and/or communication burden of MPC. However, it generally requires a priori knowledge of the closed-loop system behavior along with the communication characteristics for designing the event-trigger policy. This paper attempts to solve this challenge by proposing an efficient eMPC framework and demonstrates successful implementation of this framework on the autonomous vehicle path following. First of all, a model-free reinforcement learning (RL) agent is used to learn the optimal event-trigger policy without the need for a complete dynamical system and communication knowledge in this framework. Furthermore, techniques including prioritized experience replay (PER) buffer and long short-term memory (LSTM) are employed to foster exploration and improve training efficiency. In this paper, we use the proposed framework with three deep RL algorithms, i.e., Double Q-learning (DDQN),

Proximal Policy Optimization (PPO), and Soft Actor-Critic (SAC), to solve this problem. Results show that all three deep RL-based eMPC (deep-RL-eMPC) can achieve better evaluation performance than the conventional threshold-based and previous linear Q-based approach in the autonomous path following. In particular, PPO-eMPC with LSTM and DDQN-eMPC with PER and LSTM obtain a superior balance between the closed-loop control performance and event-trigger frequency.

Resumo: O controle preditivo de modelo acionado por evento (eMPC) é um método de controle ideal popular com o objetivo de aliviar a carga de computação e/ou comunicação do MPC. No entanto, ele geralmente requer um conhecimento prévio do comportamento do sistema em malha fechada, juntamente com as características de comunicação para projetar a política de acionamento de eventos. Este artigo tenta resolver esse desafio propondo uma estrutura eficiente de eMPC e demonstra a implementação bem-sucedida dessa estrutura no acompanhamento do caminho do veículo autônomo. Em primeiro lugar, um agente de aprendizagem por reforço (RL) sem modelo é usado para aprender a política ideal de acionamento de eventos sem a necessidade de um sistema dinâmico completo e conhecimento de comunicação nessa estrutura. Além disso, técnicas como o buffer de repetição de experiência priorizada (PER) e a memória de longo prazo (LSTM) são empregadas para promover a exploração e aumentar a eficiência do treinamento. Neste artigo, usamos a estrutura proposta com três algoritmos de RL profunda, ou seja, Double Q-learning (DDQN), Proximal Policy Optimization (PPO) e Soft Actor-Critic (SAC), para resolver esse problema. Os resultados mostram que todos os três eMPC baseados em RL profunda (deep-RL-eMPC) podem obter melhor desempenho de avaliação do que a abordagem convencional baseada em limiar e a abordagem anterior baseada em Q linear no seguimento de caminho autônomo. Em particular, o PPO-eMPC com LSTM e o DDQN-eMPC com PER e LSTM obtêm um equilíbrio superior entre o desempenho do controle de loop fechado e a frequência de disparo de eventos.

Estudo

Data de execução: 2024
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O artigo propõe um framework de controle preditivo baseado em modelo com acionamento por eventos (eMPC) utilizando aprendizado por reforço profundo (deep RL) para otimizar a política de acionamento de eventos em veículos autônomos, visando equilibrar desempenho de controle e eficiência computacional.

Hipóteses avaliadas:

- O uso de deep RL para aprender a política de acionamento de eventos em eMPC supera métodos convencionais baseados em limiares fixos.
- Técnicas como Prioritized Experience Replay (PER) e Long Short-Term Memory (LSTM) melhoram a eficiência do treinamento e o desempenho do controle.

- O parâmetro ρc na função de recompensa permite ajustar o equilíbrio entre desempenho de controle e frequência de acionamento de eventos.

Variáveis independentes:
- Algoritmos de deep RL (DDQN, PPO, SAC).
- Técnicas de aprimoramento (PER, LSTM).
- Valor do parâmetro ρc(0, 0.001, 0.01).
- Modelo dinâmico do veículo (não linear, com controle longitudinal e lateral).

Variáveis dependentes:
- **Desempenho de controle:** Erro de seguimento de trajetória (Empc).
- **Eficiência computacional:** Frequência de acionamento de eventos (Af).
- **Retorno total (R):** Combinação do desempenho e penalização por acionamento.

Participantes:
- Modelos de simulação de veículos autônomos.
- Algoritmos de RL testados (DDQN, PPO, SAC) com/sem PER e LSTM.
- Métodos comparativos: abordagem baseada em limiar [27] e Q-learning linear (LSTDQ) [42].

Material:
- Modelo dinâmico não linear de veículo (equações 1a-1f).
- Ambiente de simulação com trajetória sinusoidal (equação 5).
- Framework de treinamento RL com funções de recompensa (equação 12).
- Ferramentas: Python, bibliotecas de RL (e.g., TensorFlow/PyTorch).

Planejamento do estudo:

1. **Treinamento**:
    - Algoritmos off-policy (DDQN, SAC) treinados por 50.000 passos.
    - Algoritmo on-policy (PPO) treinado por 1.000 episódios.
    - Parâmetros: γ=0.99, batch size = 64, taxa de aprendizado = 10^−4.
2. **Avaliação**:
    - Comparação de desempenho entre algoritmos sob diferentes ρc.
    - Métricas: Empc, Af, e retorno R.
    - Análise qualitativa de padrões de acionamento (Figuras 3-5).

Ameaças à validade:

- Validade externa: Resultados baseados em simulação; não testado em cenários reais ou ambientes complexos (e.g., CARLA).
- Validade interna: Uso de um modelo específico de veículo e trajetória sinusoidal, limitando generalização.
- Viés de implementação: Dependência de hiperparâmetros (e.g., ρc) ajustados manualmente.

Resultados:

- PPO + LSTM obteve melhor desempenho para ρc=0 e 0.001 (menor Empc).
- DDQN + LSTM + PER destacou-se para ρc=0.01, reduzindo Af para 0.255 com Empc=0.171.
- Todos os métodos baseados em deep RL superaram abordagens convencionais (limiar e LSTDQ).
- O aumento de ρc reduziu a frequência de acionamento, mas aumentou Empc.

Comentários adicionais:

- Limitação: Ambiente de simulação simplificado.
- Trabalhos futuros: Testes em ambientes mais realistas (e.g., CARLA), incorporação de ruídos e validação em hardware.
- Contribuição: Framework model-free que não requer conhecimento prévio da dinâmica do sistema.

Referências relevantes

- [27]: Comparação de políticas de acionamento para eMPC.
- [42]: Trabalho anterior com Q-learning linear (base para comparação).
- [43]: RL para controle com acionamento por eventos (contexto diferente).
- [46], [47], [48]: Algoritmos de deep RL (DDQN, PPO, SAC).

Justificativa:

O estudo é relevante pois aborda o desafio de reduzir a carga computacional do MPC tradicional em veículos autônomos, combinando-o com técnicas de RL para otimizar a política de acionamento. Isso permite aplicações em sistemas com restrições de tempo real, mantendo desempenho e segurança.

12º)

Título do Artigo: Distributed Nonlinear Model Predictive Control and Reinforcement Learning

Autores: Ifrah Saeed, Tansu Alpcan, Sarah M. Erfani, M. Berkay Yilmaz

Data da Publicação: 02/11/2019

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Coordinating two or more dynamic systems such as autonomous vehicles or satellites in a distributed manner poses an important research challenge. Multiple approaches to this problem have been proposed including Nonlinear Model Predictive Control (NMPC) and its model-free counterparts in reinforcement learning (RL) literature such as Deep QNetwork (DQN). This initial study aims to compare and contrast the optimal control technique, NMPC, where the model is known, with the popular model-free RL method, DQN. Simple distributed variants of these for the specific problem of balancing and synchronising two highly unstable cart-pole systems are investigated numerically. We found that both NMPC and trained DQN work optimally under ideal model and small communication delays. While NMPC performs sub-optimally under a model-mismatch scenario, DQN performance naturally does not suffer from this. Distributed DQN needs a lot of realworld experience to be trained but once it is trained, it does not have to spend its time finding the optimal action at every time-step like NMPC. This illustrative comparison lays a foundation for hybrid approaches, which can be applied to complex multi-agent scenarios.

Resumo: A coordenação de dois ou mais sistemas dinâmicos, como veículos autônomos ou satélites, de forma distribuída, representa um importante desafio de pesquisa. Várias abordagens para esse problema foram propostas, incluindo o controle preditivo de modelo não linear (NMPC) e suas contrapartes sem modelo na literatura de aprendizagem por reforço (RL), como a Deep QNetwork (DQN). Este estudo inicial tem como objetivo comparar e contrastar a técnica de controle ideal, NMPC, em que o modelo é conhecido, com o método popular de RL sem modelo, DQN. As variantes distribuídas simples desses métodos para o problema específico de balanceamento e sincronização de dois sistemas de polo de carro altamente instáveis são investigadas numericamente. Descobrimos que tanto o NMPC quanto o DQN treinado funcionam de forma ideal sob um modelo ideal e com pequenos atrasos de comunicação. Embora o NMPC tenha um desempenho abaixo do ideal em um cenário de incompatibilidade de modelos, o desempenho do DQN naturalmente não sofre com isso. O DQN distribuído precisa de muita experiência no mundo real para ser treinado, mas, uma vez treinado, ele não precisa gastar seu tempo para encontrar a ação ideal em cada etapa, como o NMPC. Essa comparação ilustrativa estabelece uma base para abordagens híbridas, que podem ser aplicadas a cenários complexos de vários agentes.

Estudo

Data de execução: 2019
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O estudo compara o desempenho do Controle Preditivo de Modelo Não Linear (NMPC) e da Rede Deep Q-Network (DQN) em sistemas distribuídos de carrinho-pêndulo, visando equilibrar e sincronizar dois sistemas dinâmicos instáveis. O objetivo é avaliar a eficácia dessas abordagens em cenários ideais, com discrepância de modelo e sob restrições de comunicação.

Hipóteses avaliadas:

- NMPC é eficaz sob modelo ideal, mas falha em cenários de discrepância de modelo.
- DQN, apesar de exigir treinamento extensivo, adapta-se melhor a incertezas e restrições de comunicação.
- Ambos os métodos mantêm sincronização sob pequenos atrasos de comunicação, mas com comportamentos distintos.

Variáveis independentes:

- Tipo de controle (NMPC vs. DQN).
- Condições do modelo (ideal vs. discrepância de parâmetros).
- Restrições de comunicação (atrasos, perda de pacotes).

Variáveis dependentes:

- Estabilidade dos sistemas (ângulo do pêndulo, posição do carrinho).
- Sincronização entre os sistemas (diferença de posição e ângulo).
- Tempo computacional para obtenção de ações ótimas.

Participantes:
Dois sistemas simulados de carrinho-pêndulo, cada um atuando como agente controlador em um ambiente distribuído.

Material:

- Modelo matemático do sistema carrinho-pêndulo (equações dinâmicas).
- Ambiente de simulação em Python 3.6 (OpenAI Gym modificado).
- Processador Intel i7-8550U para execução das simulações.
- Redes neurais profundas (DQN) e algoritmos de otimização (NMPC).

Planejamento do estudo:

1. Configuração inicial: Simulação de dois sistemas de carrinho-pêndulo com parâmetros padrão.
2. Cenários testados:
    - Modelo ideal (parâmetros corretos).
    - Discrepância de modelo (alteração de massa e comprimento do pêndulo).
    - Comunicação com atraso (50 passos temporais) e perda de pacotes (10%).
3. Treinamento do DQN: Coleta de experiência real para aprendizado da política ótima.
4. Avaliação comparativa: Análise de estabilidade, sincronização e tempo computacional.

Ameaças à validade:

- Validade interna: O uso de parâmetros específicos de simulação pode limitar a generalização para outros sistemas dinâmicos.

- Validade externa: Resultados baseados em simulações podem não refletir cenários reais com ruídos ou dinâmicas mais complexas.
- Viés de implementação: DQN utiliza ações discretas (21 níveis), enquanto NMPC opera com ações contínuas, o que pode afetar a comparação direta.

Resultados:

- **NMPC:**
  - Ótimo em modelo ideal, mas falha em sincronização sob discrepância de parâmetros.
  - Estável sob atrasos de comunicação, mas com aumento do tempo computacional**.**
- **DQN:**
  - Requer treinamento extensivo, mas mantém estabilidade e sincronização mesmo com discrepância de modelo.
  - Ações discretas levam a oscilações menores, porém sincronização menos precisa sob atrasos.

Comentários adicionais:

- O estudo destaca a complementaridade entre NMPC e DQN, sugerindo abordagens híbridas futuras.
- A escolha de ações discretas no DQN pode ser uma limitação para sistemas que demandam controle contínuo.

Referências relevantes

1. Grune e Pannek (2017) - Fundamentos de NMPC.
2. Mnih et al. (2015) - Arquitetura DQN.
3. Mayne et al. (2011) - Técnicas robustas de NMPC (tube-based).
4. Hessel et al. (2018) - Algoritmo Rainbow para RL estável.

Justificativa:

A pesquisa busca preencher a lacuna na comparação entre métodos baseados em modelo (NMPC) e métodos sem modelo (DQN) em sistemas distribuídos não lineares. Os resultados fornecem insights para escolha de técnicas de controle em cenários com incertezas e restrições práticas, além de motivar o desenvolvimento de abordagens híbridas.

13º)

Título do Artigo: Leveraging AI for Enhanced Power Systems Control: An Introductory Study of Model-Free DRL Approaches

Autores: Yi Zhou, Liangcai Zhou, Zhehan Yi, Di Shi, Mengjie Guo

Data da Publicação: 03/07/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: The power grids nowadays are facing increasing complexity and uncertainty due to the continuously growing penetration of renewable energy sources, such as photovoltaic (PV) and wind power, as well as the emerging uncertain and dynamic nature of demand-side factors such as electric vehicles, portable energy storage, etc. To effectively manage these challenges, artificial intelligence (AI) technologies, particularly model-free deep reinforcement learning (DRL), have risen as a powerful tool for power system control. This paper presents an in-depth review of the state-of-the-art applications of model-free DRL in power system control. The review is focused on various model-free DRL approaches utilized in addressing uncertainty caused by stochastic factors from renewable generations, demand-side dynamics, and power system contingencies. Specifically, it investigates how model-free DRL techniques are employed to facilitate decision-making in the frequency control, voltage control, and optimal power flow control in the grid. The benefits, challenges, and limitations of these technologies are revealed, shedding light on recent advancements in the field and showcasing the performance of these methods across diverse power system scenarios. By synthesizing the findings from extensive research, this paper also highlights the limitations, key future research directions, and recommendations for deploying model-free DRL in power system control.

Resumo: Atualmente, as redes de energia estão enfrentando uma complexidade e uma incerteza crescentes devido à penetração cada vez maior de fontes de energia renováveis, como a energia fotovoltaica (PV) e eólica, bem como à natureza incerta e dinâmica emergente dos fatores do lado da demanda, como veículos elétricos, armazenamento portátil de energia etc. Para gerenciar esses desafios com eficácia, as tecnologias de inteligência artificial (IA), especialmente a aprendizagem por reforço profundo (DRL) sem modelos, surgiram como uma ferramenta poderosa para o controle do sistema de energia. Este artigo apresenta uma análise aprofundada das aplicações de ponta da DRL sem modelos no controle do sistema de energia. A análise se concentra em várias abordagens de DRL sem modelo utilizadas para lidar com a incerteza causada por fatores estocásticos de gerações renováveis, dinâmica do lado da demanda e contingências do sistema de energia. Especificamente, ela investiga como as técnicas de DRL sem modelos são empregadas para facilitar a tomada de decisões no controle de frequência, no controle de tensão e no controle ideal do fluxo de energia na rede. Os benefícios, desafios e limitações dessas tecnologias são revelados, lançando luz sobre os recentes avanços no campo e demonstrando o desempenho desses métodos em diversos cenários do sistema de energia. Ao sintetizar as descobertas de uma extensa pesquisa, este documento também destaca as limitações, as principais direções de pesquisas futuras e as recomendações para a implantação de DRL sem modelo no controle do sistema de energia.

Estudo

Data de execução: 2024
Local: Ambiente Computacional
Tipo: Simulação

Descrição:

O artigo realiza uma revisão abrangente das aplicações de abordagens de Deep Reinforcement Learning (DRL) sem modelo no controle de sistemas de energia modernos, focando em desafios como a integração de fontes renováveis, dinâmicas de demanda e contingências. O estudo analisa métodos de DRL para controle de tensão, frequência e fluxo de potência ótimo, destacando suas vantagens, limitações e tendências futuras.

Hipóteses avaliadas:

- Métodos de DRL sem modelo superam abordagens tradicionais em adaptabilidade, resposta em tempo real e eficiência no controle de sistemas de energia.
- Abordagens baseadas em políticas (ex: DDPG, PPO) são mais estáveis e adequadas para espaços de ação contínua.
- Métodos multiagentes (MA-DRL) permitem controle descentralizado e robusto em sistemas complexos.

Variáveis independentes:

- Algoritmos de DRL (DQN, DDPG, PPO, SAC, TD3).
- Parâmetros de treinamento (número de episódios, funções de ativação, estrutura de redes neurais).
- Cenários de teste (sistemas de transmissão, distribuição, microredes).
- Disponibilidade de dados (simulados, históricos, em tempo real).

Variáveis dependentes:

- Estabilidade de tensão e frequência.
- Minimização de perdas de energia e custos operacionais.
- Tempo de resposta a contingências.
- Robustez a incertezas (renováveis, demanda).

Participantes:

- Sistemas de energia simulados (ex: IEEE 123-bus, Illinois 200-bus).
- Redes reais (ex: Província de Liaoning, China; East China Power Grid).

Material:

- Simuladores de fluxo de carga (ex: PowerWorld, PSS/E).
- Dados de geração renovável, carga e contingências.
- Frameworks de DRL (ex: TensorFlow, PyTorch).
- Modelos de redes neurais profundas.

Planejamento do estudo:

- Revisão sistemática de artigos publicados entre 2020–2024.
- Classificação dos métodos em categorias (valor vs. política, mono vs. multiagente).
- Análise comparativa de desempenho em diferentes aplicações (tensão, frequência, OPF).
- Identificação de lacunas (ex: generalização, escalabilidade) e recomendações futuras.

Ameaças à validade:

- Validade interna: Dependência de dados simulados e simplificações em modelos de teste.
- Validade externa: Dificuldade de generalização para sistemas reais e de grande escala.
- Viés de seleção: Foco em estudos com resultados positivos para DRL.
- Recursos computacionais: Treinamento intensivo limita aplicação prática**.**

Resultados:

- Métodos baseados em políticas (DDPG, PPO) destacaram-se em controle contínuo e estabilidade.
- MA-DRL mostrou eficácia em cenários descentralizados (ex: controle de tensão em redes ativas).
- DRL reduziu desvios de tensão/frequência em até 99,92% em testes (ex: Grid Mind [20]).
- Limitações incluem necessidade de dados extensivos e falta de interpretabilidade.

Comentários adicionais:

- A integração de IA com conhecimento físico (ex: restrições operacionais) é crítica para segurança.
- Métodos human-in-the-loop (ex: HL-SAC [58]) podem mitigar riscos em implementações reais.
- Colaboração academia-indústria é essencial para padronização e adoção prática.

Referências relevantes

- [20] Grid Mind (DDPG para controle de tensão).
- [44] MA-DDPG em sistemas multiárea.
- [32] SSAC para controle seguro de frequência.
- [58] HL-SAC com intervenção humana.
- [80] GPT-DRL para OPF com interpretabilidade.

Justificativa:

A transição para sistemas de energia com alta penetração de renováveis exige métodos de controle adaptativos. O DRL sem modelo emerge como uma solução promissora, mas sua implementação requer superação de desafios como escalabilidade, segurança e interpretabilidade. Esta revisão sistematiza avanços recentes, orientando pesquisadores e profissionais na adoção responsável dessas tecnologias.

14º)

Título do Artigo: A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning

Autores: Pouria Qashqai, Mohammad Babaie, Rawad Zgheib, Kamal Al-Haddad

Data da Publicação: 22/09/2023

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: This paper presents a novel model-free switching and control method for three-level neutral point clamped (NPC) converter using deep reinforcement learning (DRL). Our approach targets two primary control objectives: voltage balancing and current control. In this method, voltage balancing, current control and selection of optimal switches are achieved using a reward function which is calculated based on various signals measured as observations of the DRL agent. Since the action space is discrete, a deep Q-network (DQN) agent is utilized. DQN is used due to its capability of handling high-dimensional state spaces. In order to highlight its pros and cons, the proposed method is compared with model predictive control (MPC), which is another popular non-linear control method for power electronic converters. The proposed method is evaluated and compared with the MPC method in grid-connected mode using simulations in Matlab/Simulink. To evaluate the practical performance of the DRL method, various experimental results are obtained. The simulation and experimental results demonstrate that the proposed method effectively achieves accurate voltage balancing and ensures steady operation even in the presence of various dynamic changes, including variations in the reference currents and grid voltage. Additionally, the method successfully handles uncertainties, such as sensor measurement noise, and accommodates parameter variations, such as changes in the capacity of the DC-link capacitors and line impedance. The results demonstrate that this method exhibits superior adaptability to real-time changes and uncertainties, delivering more robust performance compared to similar conventional methods like MPC. Thus, this method can be considered a promising approach for intelligent control of power electronic converters, especially when conventional methods such as MPC face challenges in performance and accuracy under severe parameter variations and uncertainties.

Resumo: Este artigo apresenta um novo método de controle e comutação sem modelos para o conversor de ponto neutro (NPC) de três níveis usando o aprendizado por reforço profundo (DRL). Nossa abordagem visa a dois objetivos principais de controle:

balanceamento de tensão e controle de corrente. Nesse método, o equilíbrio da tensão, o controle da corrente e a seleção dos interruptores ideais são obtidos por meio de uma função de recompensa calculada com base em vários sinais medidos como observações do agente DRL. Como o espaço de ação é discreto, é utilizado um agente de rede Q profunda (DQN). O DQN é usado devido à sua capacidade de lidar com espaços de estado de alta dimensão. Para destacar seus prós e contras, o método proposto é comparado com o controle preditivo de modelo (MPC), que é outro método popular de controle não linear para conversores eletrônicos de potência. O método proposto é avaliado e comparado com o método MPC no modo conectado à rede usando simulações no Matlab/Simulink. Para avaliar o desempenho prático do método DRL, são obtidos vários resultados experimentais. A simulação e os resultados experimentais demonstram que o método proposto alcança com eficácia o equilíbrio preciso da tensão e garante uma operação estável mesmo na presença de várias alterações dinâmicas, inclusive variações nas correntes de referência e na tensão da rede. Além disso, o método lida com sucesso com incertezas, como o ruído de medição do sensor, e acomoda variações de parâmetros, como mudanças na capacidade dos capacitores do link CC

Estudo

Data de execução: 2013
Local: Ambiente Computacional e Bancada de Laboratório
Tipo: Simulação no MATLAB e Bancada

Descrição:

O artigo propõe um método de controle e comutação sem modelo para conversores NPC (Neutral Point Clamped) de três níveis, utilizando Deep Reinforcement Learning (DRL). O método visa balanceamento de tensão e controle de corrente, comparando-se com o Model Predictive Control (MPC). Foram realizadas simulações no Matlab/Simulink e testes experimentais para validar a robustez do método em cenários dinâmicos, variações de parâmetros, incertezas e ruídos.

Hipóteses avaliadas:

1. O método DRL proposto é mais adaptável a incertezas, variações de parâmetros e ruídos em comparação com métodos convencionais como o MPC.
2. O DRL é capaz de controlar eficientemente o balanceamento de tensão e a corrente do conversor sem exigir um modelo matemático preciso.

Variáveis independentes:

- Variações nas correntes de referência (idref, iqref).
- Mudanças na tensão da rede elétrica.
- Aumento da indutância e resistência da rede.
- Degradação da capacitância dos capacitores do DC-link.
- Adição de ruído Gaussiano nas medições de sensores.

Variáveis dependentes:

- Precisão no balanceamento das tensões dos capacitores (VC1, VC2).
- Acompanhamento das correntes de referência (id, iq).
- THD (Distorção Harmônica Total) das correntes de saída.
- Frequência de comutação e perdas associadas.
- Estabilidade do sistema sob perturbações e variações dinâmicas.

Participantes:
Não se aplica (estudo de simulação e experimentos com equipamentos)**.**

Material:
- **Simulação:** Matlab/Simulink.
- **Hardware experimental:** dSPACE 1202, OP8662 (medições de tensão/corrente), placa de potência NPC, autotransformador, cargas lineares (50 mH, 40 Ω) e não lineares (80 Ω, 2200 μF).
- **Parâmetros de treinamento**: Taxa de aprendizado (0.001), tamanho do mini-batch (320), fator de desconto (0.01).

Planejamento do estudo:

1. Treinamento do agente DRL (DQN): Configuração da função de recompensa, seleção de observações (tensões, correntes) e ações (27 estados de comutação discretos).
2. Simulações: Comparação com MPC em cenários de operação em regime permanente, respostas a degraus, variações de parâmetros e ruídos.
3. Validação experimental: Testes em bancada com cargas lineares/não lineares e análise de THD, sincronismo com a rede e compensação de potência reativa.

Ameaças à validade:

- Dependência crítica do projeto da função de recompensa do DRL.
- Treinamento demorado do agente DRL.
- Comparação limitada ao MPC, sem avaliação contra outros métodos (e.g., controle deslizante).
- Restrição a cenários específicos de teste (e.g., variações de até 20% nos parâmetros).

Resultados:

- **DRL vs. MPC:** O DRL apresentou menor sensibilidade a ruídos (25 dB SNR), adaptação mais rápida a variações dinâmicas (e.g., degraus em idref) e menor frequência de comutação (2,5 kHz vs. ~10 kHz no MPC).
- **THD:** 3,62% (DRL) vs. 2,44% (MPC), porém com perdas menores no DRL.
- **Resiliência:** O DRL manteve estabilidade sob degradação de capacitores (+15%) e ruído, enquanto o MPC tornou-se instável com ruído.

Comentários adicionais:

- Destaque para a abordagem model-free, eliminando a necessidade de modelos matemáticos complexos.
- Limitação: Treinamento do DRL exige recursos computacionais significativos.
- Sugestão: Estender o método para outros conversores multinível e explorar técnicas de transfer learning para reduzir tempo de treinamento.

Referências relevantes

- [29], [30], [31]: Aplicações de DRL em conversores CC/CC.
- [40], [41]: Uso de DRL em conversores NPC.
- [10], [11], [12]: Fundamentos do MPC em eletrônica de potência.

Justificativa:

O estudo justifica-se pela necessidade de métodos de controle robustos para conversores de potência em aplicações com incertezas (e.g., energias renováveis, veículos elétricos), onde métodos tradicionais como MPC têm limitações em cenários dinâmicos e não modelados. O DRL surge como alternativa promissora por sua capacidade de generalização e adaptação em tempo real.

15º)

Título do Artigo: Reinforcement Learning-Based Predictive Control for Power Electronic Converters

Autores: Yihao Wan, Qianwen Xu, Tomislav Dragičević

Data da Publicação: 29/10/2024

Veículo de Publicação: IEEE

Fonte: IEEE Xplore

Abstract: Finite-set model predictive control (FS-MPC) appears to be a promising and effective control method for power electronic converters. Conventional FS-MPC suffers from the time-consuming process of weighting factor selection, which significantly impacts control performance. Another ongoing challenge of FS-MPC is its dependence on the prediction model for desirable control performance. To overcome the above issues, we propose to apply reinforcement learning (RL) to FS-MPC for power converters. The RL algorithm is first employed for the automatic weighting factor design of the FS-MPC, aiming to minimize the total harmonic distortion (THD) or reduce the average switching frequency. Furthermore, by formulating the incentive for the RL agent with the cost function of the predictive algorithm, the agent learns autonomously to find the optimal switching policy for the power converter by imitating the predictive controller without prior knowledge of the system model. Finally, a deployment framework that allows for experimental validation of the proposed RL-based methods on a practical FS-MPC regulated stand-alone converter configuration is presented. Two exemplary control

objectives are demonstrated to show the effectiveness of the proposed RL-aided weighting factor tuning method. Moreover, the results show a good match between the model-free RL-based controller and the FS-MPC performance.

Resumo: O controle preditivo de modelo de conjunto finito (FS-MPC) parece ser um método de controle promissor e eficaz para conversores eletrônicos de potência. O FS-MPC convencional sofre com o processo demorado de seleção do fator de ponderação, que afeta significativamente o desempenho do controle. Outro desafio constante do FS-MPC é sua dependência do modelo de previsão para obter um desempenho de controle desejável. Para superar os problemas acima, propomos a aplicação do aprendizado por reforço (RL) ao FS-MPC para conversores de energia. O algoritmo de RL é primeiramente empregado para o projeto automático do fator de ponderação do FS-MPC, com o objetivo de minimizar a distorção harmônica total (THD) ou reduzir a frequência média de comutação. Além disso, ao formular o incentivo para o agente de RL com a função de custo do algoritmo preditivo, o agente aprende de forma autônoma a encontrar a política de comutação ideal para o conversor de energia, imitando o controlador preditivo sem conhecimento prévio do modelo do sistema. Por fim, é apresentada uma estrutura de implementação que permite a validação experimental dos métodos baseados em RL propostos em uma configuração prática de conversor autônomo regulado por FS-MPC. Dois objetivos de controle exemplares são demonstrados para mostrar a eficácia do método de ajuste do fator de ponderação auxiliado por RL proposto. Além disso, os resultados mostram uma boa correspondência entre o controlador baseado em RL sem modelo e o desempenho do FS-MPC.

Estudo

Data de execução: 2024
Local: Ambiente Computacional e Bancada de Laboratório
Tipo: Simulação no MATLAB e Bancada

Descrição:

O artigo propõe o uso de Reinforcement Learning (RL) para aprimorar o Finite-Set Model Predictive Control (FS-MPC) em conversores de potência, abordando dois desafios principais: (1) a seleção automática de fatores de ponderação na função de custo do FS-MPC e (2) a redução da dependência do modelo preditivo. São apresentados dois métodos: um baseado em Deep Deterministic Policy Gradient (DDPG) para ajuste automático dos fatores de ponderação e outro baseado em Deep Q-Network (DQN) para imitar o comportamento do FS-MPC sem conhecimento prévio do modelo do sistema.

Hipóteses avaliadas:

1. O RL pode automatizar o ajuste dos fatores de ponderação do FS-MPC, melhorando o desempenho de controle (e.g., minimização de THD e frequência de chaveamento).
2. Um controlador baseado em RL pode emular o FS-MPC sem dependência de modelos matemáticos, mantendo desempenho comparável.

Variáveis independentes:

- Fatores de ponderação (λd, λsw) no FS-MPC.
- Política de seleção de estados de chaveamento (ações do agente RL).
- Configurações do ambiente de simulação e experimental (e.g., carga, tensão de referência).

Variáveis dependentes:

- Total Harmonic Distortion (THD).
- Frequência média de chaveamento (fsw).
- Desempenho transitório (e.g., resposta a variações de carga).
- Estabilidade da tensão de saída.

Participantes:

- Sistema experimental com conversor VSC (Voltage Source Converter) de dois níveis.
- Cargas lineares e não lineares (incluindo cenários de carga desbalanceada).
- Plataforma de controle IMPERIX B-Box RCP para implementação prática.

Material:

- Modelo matemático do conversor VSC em referencial αβ.
- Ambiente de simulação MATLAB/Simulink.
- Hardware experimental: módulos de potência IMPERIX PEB 8024, filtro LC, carga linear.
- Algoritmos de RL: DDPG (para ajuste de fatores) e DQN (para imitação do FS-MPC).

Planejamento do estudo:

3. Treinamento offline dos agentes RL (DDPG e DQN) em simulação, utilizando modelos matemáticos do sistema.
4. Transferência para implementação prática via plataforma IMPERIX, validando os resultados em um conversor real.
5. Casos de estudo:
    - Cenário A: Minimização de THD com DDPG.
    - Cenário B: Trade-off entre THD e fsw.
    - Comparação com métodos baseados em ANN (Artificial Neural Network).
    - Validação do controlador DQN em condições transitórias e cargas desconhecidas.

Ameaças à validade:

- Validade externa: Resultados limitados a conversores VSC de dois níveis; necessária validação em topologias mais complexas (e.g., multinível).

- Dependência de simulação: O treinamento inicial dos agentes RL foi realizado em ambiente simulado, o que pode não capturar totalmente ruídos e não linearidades do sistema real.
- Generalização: O desempenho do controlador DQN em cenários não treinados (e.g., cargas altamente não lineares) não foi explorado em profundidade.

Resultados:

- DDPG: Redução de THD de 3,37% para 2,22% (Cenário A) e THD de 2,61% com fsw reduzida para 3,62 kHz (Cenário B).
- DQN: Desempenho comparável ao FS-MPC convencional, com diferença de 0,12% em THD e 0,37 kHz em fsw.
- Resposta transitória: Recuperação rápida da tensão após degrau de carga (100% → 50% em 4 ms).
- Cargas desbalanceadas: THD de 3,56% sem conhecimento prévio da carga.

Comentários adicionais:

- O método proposto elimina a necessidade de otimização manual ou supervisionada, reduzindo custos computacionais e tempo de desenvolvimento.
- A implementação prática demonstra viabilidade, mas a complexidade de treinamento offline pode ser uma barreira para aplicações em tempo real.
- Futuros trabalhos podem explorar a extensão para conversores multinível e incorporar garantias de estabilidade (e.g., funções de Lyapunov).

Referências relevantes

- Dragicevic et al. (2016; 2021): Fundamentos de controle em microgrids e FS-MPC.
- Rodriguez e Cortes (2012): Bases do controle preditivo para conversores.
- Watkins e Dayan (1992): Algoritmo Q-learning.
- Mnih et al. (2015): Deep Q-Networks (DQN).
- Lillicrap et al. (2015): Deep Deterministic Policy Gradient (DDPG).

Justificativa:

O estudo foi conduzido para superar limitações do FS-MPC convencional, como a seleção manual de fatores de ponderação e a dependência de modelos precisos. O RL oferece uma abordagem model-free e autônoma, adequada para sistemas complexos e não lineares, além de permitir otimização multiobjetivo sem etapas adicionais de pós-processamento. A validação experimental reforça a aplicabilidade em cenários reais.

ANÁLISE DOS RESULTADOS

Agrupamento, comparação e discussão crítica dos trabalhos relacionados

- Tabulação de resultados (Dados quantitativos)

1. **O que já foi publicado?**

- **Controle de Inversores e Conversores:**
  - Estratégias de controle de corrente para VSI usando DQN, com análise de arquiteturas de DNN.
  - Controle sem modelo de conversores NPC trifásicos via DRL (DQN), validado experimentalmente.
  - Uso de PPO para ajustar controladores de retificadores multiníveis (PEC9) em carregadores de VEs.
- **Gestão de Energia em VEs e HEVs:**
  - EMS baseado em MPC-DRL para HEVs, combinando Bi-LSTM e DQN.
  - Controle térmico em VEs via DQN, minimizando consumo de energia e erros.
- **Redes Elétricas e Controle de Tensão:**
  - AWDDQN para controle de tensão em ADNs, superando DQN e DDQN.
  - RL para otimização de tarifas em estações de recarga de VEs, adaptando-se a fontes renováveis.
- **Aplicações em Veículos Autônomos:**
  - eMPC baseado em RL (PPO, SAC, DDQN) para seguimento de trajetória, com técnicas como LSTM e PER.
  - Comparação entre NMPC e DQN em sincronização de sistemas dinâmicos.

2. **Quais as teorias utilizadas por estudo?**

- **Aprendizado por Reforço (RL):**
  - Algoritmos como DQN, DDPG, AWDDQN, PPO, SAC.
  - Processos de Decisão de Markov (MDP) para modelagem de recompensas.
- **Arquiteturas de Redes Neurais:**
  - DNNs, Bi-LSTM (para previsão de velocidade), redes LSTM (em eMPC).
- **Controle Preditivo (MPC):**
  - Integração com RL (MPC-DRL) para otimização de energia em HEVs.
- **Teoria de Controle Clássico:**
  - Controle de retrocesso (BSC) ajustado via PPO para retificadores.

3. **O que é conhecido até recentemente sobre o tema?**

- **Eficácia do RL em Sistemas Não Lineares:**
  - RL lida com incertezas e dinâmicas complexas (ex: inversores, redes ADN).
  - DDPG supera DQN em controle contínuo (ex: torque em HEVs).
- **Problemas em Métodos Tradicionais:**
  - Superestimação de recompensas no DQN, resolvida com AWDDQN.
  - Dificuldade de controle em cenários de incompatibilidade de modelos (ex: NMPC vs DQN).
- **Integração com Técnicas Clássicas:**
  - Combinação de MPC com RL (MPC-DRL) para planejamento de longo prazo.

- Indicadores/Métricas/Critérios e seus valores

Métrica	Contexto de Aplicação	Valores Reportados
RMSE	Controle de corrente em inversores (DQN)	0,83 A (DQN), 0,78 A (DDPG)
THD	Qualidade da corrente em inversores	5,29% (DQN a 10 kHz), 3,17% (DDPG a 6 kHz)
Convergência	Treinamento de DDPG vs DQN em HEVs	DDPG converge mais rápido e robustamente
THD em Redes	Controle de tensão (AWDDQN)	Menor volatilidade vs DQN/DDQN
Custo Computacional	Arquiteturas de DNN em VSI	1 camada e 5 neurônios foi ideal

- Comparações críticas

- **DQN vs DDPG:**
    - DQN limita-se a ações discretas; DDPG é superior em controle contínuo (ex: HEVs).
    - DDPG reduz RMSE e THD em inversores comparado a DQN.
- **AWDDQN vs DQN/DDQN:**
    - AWDDQN resolve superestimação (DQN) e subestimação (DDQN) em redes ADN.
- **PPO vs Controladores Baseados em Regras:**
    - PPO ajusta parâmetros de BSC em retificadores, lidando melhor com cargas desbalanceadas.
- **NMPC vs DQN Distribuído:**
    - NMPC sofre com incompatibilidade de modelos; DQN não, mas requer mais dados de treinamento.

- Lacunas existentes

- **Validação Prática Limitada:**
    - Muitos estudos validados apenas em simulação (ex: IEEE 33/123-bus), com poucas implementações reais.
- **Complexidade de Treinamento:**
    - DQN distribuído requer grande volume de dados (ex: sincronização de sistemas dinâmicos).
- **Adaptação a Cenários Dinâmicos:**
    - Falta de análise em condições extremas (ex: falhas abruptas em redes ou veículos).
- **Trade-off Desempenho-Custo:**
    - Arquiteturas simples de DNN (ex: 1 camada) são preferidas, mas podem limitar aplicações complexas.
- **Integração com Fontes Renováveis:**
    - Poucos estudos exploram RL em cenários híbridos (ex: solar + eólica + VEs) com alta variabilidade.

REFERÊNCIAS

MENENDEZ, O.; RUIZ, F.; PESANTEZ, D.; VASCONEZ, J.; RODRIGUEZ, J. Model-free Neural Network-based Current Control for Voltage Source Inverter. *2024 IEEE International Conference on Automation/XXVI Congress of the Chilean Association of Automatic Control (ICA-ACCA)*, Santiago, Chile, 2024. p. 1-6. DOI: 10.1109/ICA-ACCA62622.2024.10766747.

CHEN, F. et al. Reinforcement Learning-Based Energy Management Control Strategy of Hybrid Electric Vehicles. *2022 8th International Conference on Control, Automation and Robotics (ICCAR)*, Xiamen, China, 2022. p. 248-252. DOI: 10.1109/ICCAR55106.2022.9782662.

WANG, Y. et al. Intelligent Voltage Control Method in Active Distribution Networks Based on Averaged Weighted Double Deep Q-network Algorithm. *Journal of Modern Power Systems and Clean Energy*, v. 11, n. 1, p. 132-143, jan. 2023. DOI: 10.35833/MPCE.2022.000146.

LIN, J. Voltage Regulation in Active Distribution Network with Multiagent Deep Q-Learning Approach. *2024 43rd Chinese Control Conference (CCC)*, Kunming, China, 2024. p. 7234-7238. DOI: 10.23919/CCC63176.2024.10662011.

JAIN, P. et al. Dynamic Tariff Optimization for EV Charging Stations Using Reinforcement Learning. *2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, 2024. p. 205-210. DOI: 10.1109/ICPEICES62430.2024.10719114.

PODDUBNYY, A.; NGUYEN, P.; SLOOTWEG, H. Network-Aware Online Charge Control with Reinforcement Learning. *2022 International Conference on Smart Energy Systems and Technologies (SEST)*, Eindhoven, Netherlands, 2022. p. 1-6. DOI: 10.1109/SEST53650.2022.9898460.

ZHANG, C. et al. Energy management of hybrid electric vehicles based on model predictive control and deep reinforcement learning. *2022 41st Chinese Control Conference (CCC)*, Hefei, China, 2022. p. 5441-5446. DOI: 10.23919/CCC55666.2022.9902409.

CHOI, W. et al. Reinforcement Learning-based Controller for Thermal Management System of Electric Vehicles. *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Merced, CA, USA, 2022. p. 1-5. DOI: 10.1109/VPPC55846.2022.10003470.

MENÉNDEZ, O. et al. Assessment of Deep Reinforcement Learning Algorithms for Three-Phase Inverter Control. *2023 IEEE 8th Southern Power Electronics Conference and 17th Brazilian Power Electronics Conference (SPEC/COBEP)*, Florianopolis, Brazil, 2023. p. 1-8. DOI: 10.1109/SPEC56436.2023.10407331.

GHEISARNEJAD, M. et al. Data-Driven Switching Control Technique Based on Deep Reinforcement Learning for Packed E-Cell as Smart EV Charger. *IEEE Transactions on*

Transportation Electrification, v. 11, n. 1, p. 3194-3203, fev. 2025. DOI: 10.1109/TTE.2024.3435763.

DANG, F. et al. Event-Triggered Model Predictive Control With Deep Reinforcement Learning for Autonomous Driving. *IEEE Transactions on Intelligent Vehicles*, v. 9, n. 1, p. 459-468, jan. 2024. DOI: 10.1109/TIV.2023.3329785.

SAEED, I.; ALPCAN, T.; ERFANI, S. M.; YILMAZ, M. B. Distributed Nonlinear Model Predictive Control and Reinforcement Learning. *2019 Australian & New Zealand Control Conference (ANZCC)*, Auckland, New Zealand, 2019. p. 1-3. DOI: 10.1109/ANZCC47194.2019.8945719.

ZHOU, Y. et al. Leveraging AI for Enhanced Power Systems Control: An Introductory Study of Model-Free DRL Approaches. *IEEE Access*, v. 12, p. 98189-98206, 2024. DOI: 10.1109/ACCESS.2024.3422411.

QASHQAI, P. et al. A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning. *IEEE Access*, v. 11, p. 105394-105409, 2023. DOI: 10.1109/ACCESS.2023.3318264.

WAN, Y.; XU, Q.; DRAGIČEVIĆ, T. Reinforcement Learning-Based Predictive Control for Power Electronic Converters. *IEEE Transactions on Industrial Electronics*, 2024. DOI: 10.1109/TIE.2024.3472299.