

Heurística baseada em aprendizagem por reforço para o planejamento estático da expansão da transmissão de sistemas elétricos

Antônio Matheus Araújo Villarim, Isis Didier Lins, Márcio José das Chagas Moura

Universidade Federal de Pernambuco

Av. da Arquitetura, s/n – Cidade Universitária, Recife - PE, Brasil, CEP: 50740-550

{antonio.villarim, isis.lins, marcio.cmoura}@ufpe.br

RESUMO

O planejamento da expansão da transmissão de sistemas elétricos de potência, de modo geral, é um problema de otimização que busca identificar o conjunto de reforços para a rede elétrica de transmissão com base em características econômicas e operacionais desejadas. Recentemente, observa-se um crescente interesse pela aplicação de modelos que considerem o conjunto de restrições associado ao fluxo de potência em corrente alternada, devido à maior qualidade das soluções obtidas. Entretanto, a aplicação direta dessas restrições resulta em um problema de programação não-linear inteira mista não convexo, dificultando a aplicação de métodos exatos. Motivado por publicações recentes integrando aprendizagem por reforço a problemas de otimização combinatória, este trabalho investiga uma abordagem heurística baseada na combinação da aprendizagem por reforço com um *solver* não-linear. Testes realizados em uma instância de 6 barras atingiram objetivos ótimos globais ou aproximaram-se de resultados encontrados na literatura.

PALAVRAS CHAVE. Planejamento da expansão da transmissão, aprendizagem por reforço, programação não-linear inteira mista.

ABSTRACT

Transmission expansion planning of electrical power systems is, in general, an optimization problem that seeks to identify the set of reinforcements for the electrical transmission network based on desired economic and operational characteristics. Recently, there has been growing interest in applying models that incorporate the set of constraints associated with alternating current power flow, due to the higher quality of the resulting solutions. However, the direct inclusion of these constraints results in a non-convex mixed integer non-linear programming problem, which poses challenges for the application of exact methods. Motivated by recent publications that integrate reinforcement learning into combinatorial optimization problems, this work investigates a heuristic approach based on combining reinforcement learning with a nonlinear *solver*. Tests carried out on a 6-bus system either reached global optimal objectives or closely approximated results found in the literature.

KEYWORDS. Transmission expansion planning, reinforcement learning, mixed integer non-linear programming.

1. Introdução

O planejamento da expansão da transmissão de sistemas elétricos é um problema amplamente estudado nas últimas décadas, cujo propósito consiste em determinar o momento e a localização da instalação de novos equipamentos à rede, tipicamente linhas de transmissão e transformadores, considerando restrições de natureza física, financeira e de qualidade de serviço [Gomes e Saraiva, 2019].

Tradicionalmente, estudos associados ao planejamento da expansão da transmissão empregam uma aproximação das equações associadas ao fluxo de potência em corrente alternada, denominada fluxo de potência em corrente contínua [Romero et al., 2002]. Conforme discutido em Skolfield e Escobedo [2022], essa estratégia permite que problemas de programação não-linear inteira mista possam ser aproximados por problemas de programação linear inteira mista. Entretanto, estudos como o realizado por [Mehrtash et al., 2023] evidenciam que a aplicação de modelos utilizando essas equações pode resultar em soluções substancialmente diferentes da solução de um modelo baseado no fluxo de potência em corrente alternada. Adicionalmente, [Morquecho et al., 2023] observam limitações em outras estratégias de linearização e relaxação convexa em propostas da literatura e, devido à característica não-linear e não convexa associada a esse conjunto de equações, argumentam que a utilização de meta-heurísticas ainda pode ser adequada para a obtenção de soluções viáveis em modelos compostos por um conjunto de restrições mais complexas.

Seguindo uma linha paralela, é possível ressaltar que em trabalhos como [Bengio et al., 2021], os seus autores argumentam que existe um grande potencial na integração da aprendizagem de máquina com métodos para solução de problemas de otimização combinatória, tanto através de abordagens heurísticas, como também por meio da sua integração em métodos exatos, possivelmente permitindo o avanço do estado da arte neste campo de estudos. Particularmente, [Mazyavkina et al., 2021] identificam a aprendizagem por reforço como uma metodologia promissora nesse contexto, apresentando o potencial de superar o desempenho de algoritmos existentes e automatizar o processo de construção de heurísticas.

Motivado pelos desafios associados à solução do problema do planejamento da expansão da transmissão considerando o fluxo de potência em corrente alternada, este trabalho tem como propósito investigar a viabilidade da utilização do aprendizado por reforço como método heurístico para a solução de problemas de programação não-linear inteira mista em combinação com um *solver* não-linear.

1.1. Trabalhos relacionados

Considerando de modo mais amplo a literatura associada à aplicação da aprendizagem por reforço para a solução de problemas de otimização combinatória, dois artigos recentes podem ser destacados. [Bello et al., 2016] propõem a combinação de redes neurais recorrentes com métodos de gradiente de política para resolver o problema do caixeiro viajante, obtendo resultados próximos aos ótimos para instâncias contendo até 100 cidades. Já em [Kool et al., 2018], uma estratégia utilizando camadas de atenção e o algoritmo REINFORCE é aplicada em problemas como o roteamento de veículos, onde resultados próximos aos ótimos também são observados em instâncias de até 100 cidades.

Apesar de a aplicação de métodos baseados em aprendizagem por reforço ao problema do planejamento da expansão da transmissão ainda ser um tema pouco explorado na literatura, é possível ressaltar alguns artigos recentes que investigam essa abordagem em diferentes contextos. Em [Wang et al., 2021a], os autores propõem a utilização de uma *deep Q-network* para a solução do planejamento dinâmico da expansão da transmissão, considerando um modelo de otimização de objetivo único que inclui avaliações de severidade por meio de simulações de Monte Carlo. No trabalho proposto por [Varbella et al., 2025], *deep Q-networks*, redes neurais de grafos e um

modelo de falhas em cascata em corrente alternada são combinados para solucionar um problema de otimização multiobjetivo envolvendo a relação entre custo e redução de risco causada pela expansão da rede. Já em [Wang et al., 2021b], uma *double deep Q-network* é proposta para a solução de uma formulação multiobjetivo do problema do planejamento da expansão da transmissão, considerando fatores econômicos e operacionais, incluindo a consideração de incertezas associadas à geração eólica.

Diferentemente dos outros trabalhos apresentados sobre esse tema, que propõem modelos e técnicas de simulação para avaliar cenários e conjuntos de restrições mais complexos, este trabalho pretende investigar, seguindo uma direção mais simples, a possibilidade de separar o processo de seleção das variáveis discretas do restante das variáveis contínuas em um problema de programação matemática não-linear mono-objetivo, garantindo a geração de soluções viáveis e permitindo sua adaptabilidade para outros conjuntos de restrições.

2. Metodologia

2.1. Modelo de otimização

Este trabalho adota o modelo de programação não-linear inteira mista proposto em [Rider et al., 2007] para o planejamento estático da expansão da transmissão. Seu objetivo consiste em determinar o conjunto de reforços para a rede de transmissão de menor custo total capaz de atender restrições operacionais com base em um cenário futuro para a geração e demanda do sistema. Especificamente, ressalta-se que suas restrições consideram o modelo de fluxo de potência em corrente alternada.

Objetivo:

$$\min \nu = \sum_{i \in L} c_i \cdot n_i \quad (1)$$

Sujeito a:

$$P_i^g - P_i^d = V_i \sum_{j \in N} V_j (G_{ij}(n) \cos(\theta_i - \theta_j) + B_{ij}(n) \sin(\theta_i - \theta_j)) \quad \forall i \in N, \quad (2)$$

$$Q_i^g - Q_i^d = V_i \sum_{j \in N} V_j (G_{ij}(n) \sin(\theta_i - \theta_j) - B_{ij}(n) \cos(\theta_i - \theta_j)) \quad \forall i \in N, \quad (3)$$

$$P_i^{g, \min} \leq P_i^g \leq P_i^{g, \max} \quad \forall i \in G, \quad (4)$$

$$Q_i^{g, \min} \leq Q_i^g \leq Q_i^{g, \max} \quad \forall i \in G, \quad (5)$$

$$V_i^{\min} \leq V_i \leq V_i^{\max} \quad \forall i \in N, \quad (6)$$

$$(n_i + n_i^0) S_i^{\text{de}} \leq (n_i + n_i^0) S_i^{\max} \quad \forall i \in L, \quad (7)$$

$$(n_i + n_i^0) S_i^{\text{para}} \leq (n_i + n_i^0) S_i^{\max} \quad \forall i \in L, \quad (8)$$

$$0 \leq n_i \leq n_i^{\max} \quad \forall i \in L, \quad (9)$$

$$n \in \mathbb{N}_0, P^g, Q^g, V, \theta \in \mathbb{R}. \quad (10)$$

N é o conjunto de barramentos do sistema. L é o conjunto de linhas existentes e candidatas. G é um subconjunto de N , representando os nós nos quais existe uma conexão com uma unidade de geração a um barramento existente. As variáveis de decisão do problema são n , um vetor de valores inteiros que representa a quantidade de linhas paralelas construídas dentre os possíveis candidatos, P_i^g , Q_i^g , V_i e θ_i , respectivamente representando, para cada barramento i : a injeção de potência ativa por parte da unidade geradora, a injeção de potência reativa por parte da unidade geradora, a tensão e o ângulo de fase.

As restrições 2 e 3 representam as equações associadas ao fluxo de potência em corrente alternada. Os limites associados à potência ativa e reativa dos geradores são determinados pelas restrições 4 e 5, enquanto que os limites de tensão por barramento são determinados pela restrição 6. As restrições 7 e 8 garantem que o fluxo de potência através de cada linha de transmissão ou transformador não exceda sua capacidade máxima. Por fim, a restrição 9 estabelece o número máximo de elementos paralelos por conexão candidata.

Os elementos $G_{ij}(n)$ e $B_{ij}(n)$ representam as partes reais e imaginárias da matriz admitância $Y(n)$ da rede em estudo em função do vetor de decisão n . Assumindo que todos os elementos paralelos são idênticos, seus elementos podem ser obtidos a partir da resistência série r_{ij} , reatância série x_{ij} e susceptância em derivação b_{ij}^{sh} , além da quantidade de conexões presentes na rede atual n_{ij}^0 e no plano de expansão n_{ij} , ambos os valores indexados pelo par de barramentos que conectam:

$$y_{ij} = \frac{1}{r_{ij} + j \cdot x_{ij}} = g_{ij} - jb_{ij} \quad (11)$$

$$Y_{ij}(n) = \begin{cases} -(n_{ij} + n_{ij}^0) \cdot y_{ij}, & \text{se } i \neq j \\ \sum_{j \in N} (n_{ij} + n_{ij}^0)(y_{ij} + 0.5 \cdot j \cdot b_{ij}^{sh}), & \text{se } i = j \end{cases} \quad (12)$$

$$G_{ij} = \text{Re}(Y_{ij}) \quad (13)$$

$$B_{ij} = \text{Im}(Y_{ij}) \quad (14)$$

Os termos S_i^{de} e S_i^{para} representam a injeção de potência aparente em cada um dos terminais das linhas de transmissão. Conforme apresentado em [Rider et al., 2007], seus valores podem ser obtidos por meio do seguinte conjunto de equações:

$$S_{ij}^{\text{de}} = \sqrt{(P_{ij}^{\text{de}})^2 + (Q_{ij}^{\text{de}})^2} \quad (15)$$

$$P_{ij}^{\text{de}} = V_i^2 g_{ij} - V_i V_j (g_{ij} \cos \theta_{ij} + b_{ij} \sin \theta_{ij}) \quad (16)$$

$$Q_{ij}^{\text{de}} = -V_i^2 (b_{ij}^{sh} + b_{ij}) - V_i V_j (g_{ij} \sin \theta_{ij} - b_{ij} \cos \theta_{ij}) \quad (17)$$

$$S_{ij}^{\text{para}} = \sqrt{(P_{ij}^{\text{para}})^2 + (Q_{ij}^{\text{para}})^2} \quad (18)$$

$$P_{ij}^{\text{para}} = V_j^2 g_{ij} - V_i V_j (g_{ij} \cos \theta_{ij} - b_{ij} \sin \theta_{ij}) \quad (19)$$

$$Q_{ij}^{\text{para}} = -V_j^2 (b_{ij}^{sh} + b_{ij}) + V_i V_j (g_{ij} \sin \theta_{ij} + b_{ij} \cos \theta_{ij}) \quad (20)$$

2.2. Processo de decisão de Markov

Com o propósito de aliviar o esforço computacional associado à solução de problemas de programação não-linear inteira mista, propõe-se uma estratégia heurística composta por uma etapa de modificação das variáveis discretas conduzida por um algoritmo de aprendizagem por reforço, sucedida por uma etapa de verificação da viabilidade do problema associado às suas restrições contínuas, efetuada por um *solver* de programação não-linear. É importante destacar que essa abordagem é encontrada de forma similar em outros trabalhos como [Rider et al., 2007], que propõe uma heurística construtiva para o mesmo problema.

Para aplicação do aprendizado por reforço nesse contexto, é necessário formular o processo de decisão de Markov associado. Em primeiro lugar, seu espaço de estados consiste em todas as possíveis combinações de reforços que podem ser incluídas na rede, no qual cada estado é modelado como um vetor de valores inteiros limitados conforme a restrição 9, representando a

quantidade de elementos paralelos utilizados em cada opção considerada. O espaço de ações, por sua vez, assume uma característica construtiva, visto que consiste nos incrementos unitários de reforços, dentre as opções possíveis, a serem adicionados ao plano de cada vez, considerando que o algoritmo sempre é iniciado com um plano vazio. O processo construtivo é determinado como concluído quando uma ação resulta em um estado válido que retorna uma solução viável do *solver*. Por sua vez, a recompensa adotada para cada estado s assume o formato conforme a expressão abaixo:

$$R(s) = \begin{cases} \exp[-(s^t \cdot \|c\|_2)], & \text{se } s \text{ é um estado válido e viável,} \\ -\alpha \cdot \exp[(s^t \cdot \|c\|_2)], & \text{se } s \text{ é um estado válido e inviável,} \\ -10, & \text{se } s \text{ é um estado inválido.} \end{cases} \quad (21)$$

$\|c\|_2$ representa a norma euclidiana do vetor de custos e α é considerado como um parâmetro ajustável. Resolver um processo de decisão de Markov consiste em determinar a política que maximiza a soma esperada acumulada e descontada das recompensas [Mazyavkina et al., 2021], representada pela expressão:

$$\pi^* = \operatorname{argmax}_{\pi} \mathbb{E} \left[\sum_{t=0}^x \gamma^t R(s_t, a_t) \right] \quad (22)$$

Logo, justifica-se o uso do custo negativo na recompensa de estados válidos e viáveis como uma forma de adequar o problema de minimização ao contexto da aprendizagem por reforço. Por outro lado, a penalidade aplicada aos estados válidos e inviáveis tem como papel apresentar informações sobre o custo do plano ao longo das etapas construtivas, aplicando uma maior penalização para estados de custo mais elevado. Nos dois casos, a transformação exponencial dos custos foi utilizada como uma forma de obter uma escala de recompensas que garantam o treinamento estável do modelo. Testes foram inicialmente realizados considerando recompensas nulas para os estados válidos e inviáveis, porém, observou-se uma redução no desempenho do método para tarefas que exigissem sequências de ações mais longas. α foi definido como 10^{-3} para os experimentos realizados neste trabalho.

2.3. Método de aprendizagem por reforço

O algoritmo *Advantage Actor Critic* [Mnih et al., 2016] foi o método selecionado para resolver o processo de decisão de Markov apresentado. Seu princípio de funcionamento consiste em utilizar um modelo que representa uma política estocástica $\pi(a_t|s_t; \theta)$ que descreve uma distribuição de probabilidades sobre qual ação tomar dado um estado atual e outro modelo para estimar a função valor $V(s_t; \theta_v)$ de um determinado estado. A atualização dos modelos, conforme [Mnih et al., 2016], é realizada por meio da aplicação do gradiente:

$$\nabla_{\theta'} \log \pi(a_t|s_t; \theta') A(s_t, a_t; \theta, \theta_v) \quad (23)$$

$$A(s_t, a_t; \theta, \theta_v) = \sum_{i=0}^{k-1} \gamma^i r_{t+i} + \gamma^k V(s_{t+k}; \theta_v) \quad (24)$$

3. Experimentos e resultados preliminares

Como forma de verificar o funcionamento da metodologia proposta, o algoritmo foi aplicado na instância inicialmente introduzida por [Garver, 1970], porém considerando as modificações apresentadas em [Rider et al., 2007]. Esse sistema possui 6 barras e uma rede inicial com 6 linhas de transmissão. Nela, todas as conexões entre pares de barramentos são consideradas como candidatas para o plano de expansão, totalizando 15 possíveis conexões, nas quais todas permitem um limite

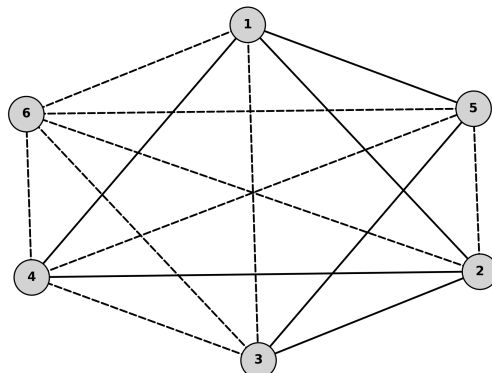


Figura 1: Representação da instância de 6 barras de [Garver, 1970] como um grafo não-dirigido. Cada aresta permite um total de até 5 conexões paralelas.

máximo de 5 linhas paralelas. A Figura 1 ilustra esta instância de modo simplificado como um grafo não-dirigido. Dois cenários foram considerados: a aplicação direta do modelo de programação matemática apresentado, considerando as conexões existentes, e a aplicação do modelo considerando uma rede inicial sem conexões.

A implementação do algoritmo *Advantage Actor Critic* (A2C) da biblioteca *Stable Baselines3* [Raffin et al., 2021] foi utilizada, aplicando redes neurais *feedforward* tanto para o modelo da política estocástica como para o aproximador da função valor. Já o *solver* não-linear selecionado foi o *Ipopt* [Wächter e Biegler, 2006], que é baseado no método de pontos interiores. Com o propósito de avaliar o desempenho e a robustez do método proposto, 10 execuções independentes do algoritmo foram realizadas em cada caso, cada uma limitada a um total de 20.000 passos. As Figuras 2 e 3 apresentam a evolução do custo do plano por episódio ao longo do processo de treinamento para os dois casos avaliados. A diferença entre seus comprimentos se deve ao fato de o procedimento de treinamento ser limitado pelo número de passos. Caso o algoritmo apresente convergência para um plano que contém mais linhas, mais passos serão necessários para a conclusão de um episódio.

No primeiro cenário, 7 das 10 tentativas apresentaram convergência para um plano de custo igual a $\$160 \cdot 10^6$, que conforme comentado em [Morquecho et al., 2023], é o ótimo global dessa instância. Já no segundo cenário, 7 dos 10 testes realizados apresentaram uma política resultando em um plano com custo igual a $\$270 \cdot 10^6$, $\$10 \cdot 10^6$ a mais do que o resultado apresentado em [Rider et al., 2007]. Entretanto, é interessante ressaltar que durante o seu treinamento, 4 execuções do algoritmo identificaram um plano com custo igual a $\$250 \cdot 10^6$, apesar de não convergirem para este valor.

4. Discussão

De modo geral, é possível verificar, com base na instância e modelo testados, que a aprendizagem por reforço apresenta potencial para aplicação como estratégia heurística na solução de problemas de programação não-linear inteira mista de objetivo único. Verifica-se que o modelo foi capaz de atingir boas soluções sem informações prévias sobre a natureza do problema, identificando boas soluções através de um procedimento de tentativa e erro com base na estrutura das recompensas e na condição de viabilidade retornada pelo *solver*.

Nota-se que o método apresentou uma característica razoavelmente robusta no primeiro cenário testado, convergindo para o ótimo global em 7 das 10 execuções em torno de 1000 episódios.

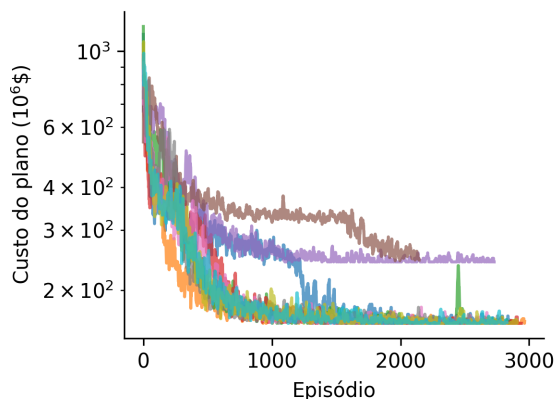


Figura 2: Curvas suavizadas para a evolução do custo do plano por episódio para cada uma das 10 execuções considerando o cenário de expansão da rede existente

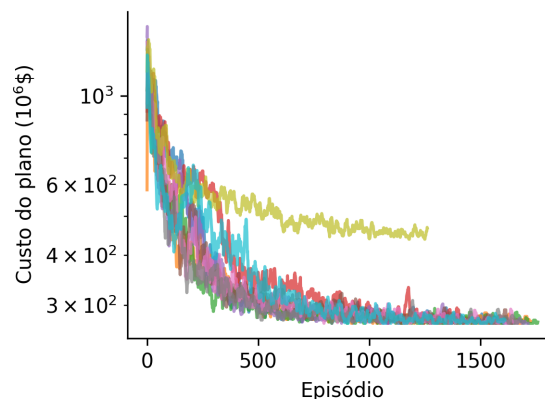


Figura 3: Curvas suavizadas para a evolução do custo do plano por episódio para cada uma das 10 execuções considerando o cenário de expansão sem uma rede prévia

Entretanto, também foi possível identificar que o modelo apresentou a tendência de convergir para soluções potencialmente ótimas locais em múltiplos testes. Apesar de a Figura 2 evidenciar que o modelo foi capaz de escapar de alguns desses estados, observa-se que o processo é geralmente lento. Com base nesse ponto, evidencia-se a necessidade de investigar mecanismos que contribuam com a exploração do modelo ao longo de seu treinamento, evitando a ocorrência deste fenômeno.

Quando comparado com outras propostas heurísticas e meta-heurísticas da literatura, é possível constatar que o método apresentado neste trabalho requer um número consideravelmente mais elevado de execuções do *solver* não-linear do que outras abordagens que também realizam a verificação das restrições do problema de modo semelhante. Por exemplo, a heurística construtiva de [Rider et al., 2007] foi capaz de atingir soluções equiparáveis com 10 verificações do *solver* no primeiro caso e 16 verificações no segundo. Por outro lado, é importante destacar que o método baseado em aprendizagem por reforço não explora nenhuma característica particular do sistema, tornando-o mais flexível para aplicação em modelos com diferentes objetivos e conjuntos de restrições.

Apesar de soluções de boa qualidade terem sido obtidas em cerca de alguns minutos na instância apresentada, é importante ressaltar que o método pode se tornar mais lento na medida em que seu conjunto de restrições se torne mais complexo e a instância apresente maior dimensionalidade. Logo, a avaliação de características associadas à escalabilidade é pertinente para melhor compreender a viabilidade dessa estratégia em contextos práticos. Uma outra direção de pesquisa interessante consiste em investigar abordagens baseadas em modelo, conforme discutido em [Mazyavkina et al., 2021].

5. Conclusão

Este trabalho apresentou uma abordagem heurística simples para a obtenção de soluções viáveis em problemas de programação não-linear inteira mista utilizando um modelo de aprendizagem por reforço para a seleção das variáveis discretas através de um procedimento construtivo, combinado com a determinação de variáveis contínuas e verificação da viabilidade da solução por meio de um *solver* não-linear. Particularmente, o problema do planejamento da expansão da transmissão considerando o fluxo de potência em corrente alternada foi estudado. Testes realizados em uma instância de 6 barras demonstraram que o método foi capaz de atingir soluções ótimas ou

próximas de referências da literatura na maioria das vezes.

Com base no exposto, tenciona-se investigar em trabalhos futuros o desempenho do método proposto em instâncias maiores e em modelos mais complexos, comparar o desempenho do método com outras abordagens exatas, heurísticas e meta-heurísticas de estado da arte e avaliar a aplicação do método em outros problemas de programação não-linear inteira mista, potencialmente considerando adaptações na função de recompensa. O avanço no desenvolvimento do método apresentado pode apresentar o potencial para a obtenção de uma heurística capaz de se adequar aos desafios mais modernos relacionados ao planejamento da expansão da transmissão.

6. Agradecimentos

Os autores agradecem ao Programa de Recursos Humanos (PRH 38.1) intitulado “Análise de Risco e Modelagem Ambiental na Exploração, Desenvolvimento e Produção de Petróleo e Gás”, financiado pela Agência Nacional de Petróleo (ANP) e gerido pela FAPESP: 24/10387-4, ‘Conselho Nacional de Desenvolvimento Científico e Tecnológico’ (CNPq): 310892/2022-8, 308924/2021-5, e ‘Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)’ - Código Financeiro 001.

Referências

- Bello, I., Pham, H., Le, Q. V., Norouzi, M., e Bengio, S. (2016). Neural combinatorial optimization with reinforcement learning. *arXiv preprint arXiv:1611.09940*.
- Bengio, Y., Lodi, A., e Prouvost, A. (2021). Machine learning for combinatorial optimization: a methodological tour d’horizon. *European Journal of Operational Research*, 290(2):405–421.
- Garver, L. L. (1970). Transmission network estimation using linear programming. *IEEE Transactions on power apparatus and systems*, (7):1688–1697.
- Gomes, P. V. e Saraiva, J. T. (2019). State-of-the-art of transmission expansion planning: A survey from restructuring to renewable and distributed electricity markets. *International Journal of Electrical Power & Energy Systems*, 111:411–424.
- Kool, W., Van Hoof, H., e Welling, M. (2018). Attention, learn to solve routing problems! *arXiv preprint arXiv:1803.08475*.
- Mazyavkina, N., Sviridov, S., Ivanov, S., e Burnaev, E. (2021). Reinforcement learning for combinatorial optimization: A survey. *Computers & Operations Research*, 134:105400.
- Mehrtash, M., Hobbs, B. F., Mahroo, R., e Cao, Y. (2023). Does choice of power flow representation matter in transmission expansion optimization? a quantitative comparison for a large-scale test system. *IEEE Transactions on Industry Applications*, 60(1):1433–1441.
- Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., e Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, p. 1928–1937. PmLR.
- Morquecho, E. G., Torres, S. P., Astudillo-Salinas, F., Ergun, H., Van Hertem, D., Castro, C. A., e Blum, C. (2023). Comparison of an improved metaheuristic and mathematical optimization based methods to solve the static AC TNEP problem. *IEEE Transactions on Power Systems*, 39(2):3240–3256.

- Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., e Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8. URL <http://jmlr.org/papers/v22/20-1364.html>.
- Rider, M., Garcia, A., e Romero, R. (2007). Power system transmission network expansion planning using AC model. *IET Generation, Transmission & Distribution*, 1(5):731–742.
- Romero, R., Monticelli, A., Garcia, A., e Haffner, S. (2002). Test systems and mathematical models for transmission network expansion planning. *IEE Proceedings-Generation, Transmission and Distribution*, 149(1):27–36.
- Skolfield, J. K. e Escobedo, A. R. (2022). Operations research in optimal power flow: A guide to recent and emerging methodologies and applications. *European Journal of Operational Research*, 300(2):387–404.
- Varbella, A., Gjorgiev, B., Sartore, F., Zio, E., e Sansavini, G. (2025). Goal-oriented graph generation for transmission expansion planning. *Engineering Applications of Artificial Intelligence*, 149:110350.
- Wächter, A. e Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106:25–57.
- Wang, Y., Chen, L., Zhou, H., Zhou, X., Zheng, Z., Zeng, Q., Jiang, L., e Lu, L. (2021a). Flexible transmission network expansion planning based on DQN algorithm. *Energies*, 14(7):1944.
- Wang, Y., Zhou, X., Shi, Y., Zheng, Z., Zeng, Q., Chen, L., Xiang, B., e Huang, R. (2021b). Transmission network expansion planning considering wind power and load uncertainties based on multi-agent DDQN. *Energies*, 14(19):6073.