

Otimização Logística em Redes de Unidades Modulares Autônomas via *Graph Neural Networks*

Otávio Trierweiler Willadino* e Jorge Otávio Trierweiler

Grupo de Intensificação, Modelagem, Simulação, Controle e Otimização de Processos (GIMSCOP)
Departamento de Engenharia Química, Universidade Federal do Rio Grande do Sul (UFRGS)

*otavio.willadino@gmail.com

RESUMO

O presente trabalho propõe uma abordagem baseada em *Graph Neural Networks* (GNNs), combinada com aprendizado por reforço profundo, para a otimização logística do Problema de Roteamento Veicular em um contexto dinâmico e descentralizado, inspirada nas Redes de Unidades Modulares Autônomas/Automatizadas, Intensificadas e Sustentáveis (RU+). O problema é modelado como um grafo, no qual os NÓS representam unidades produtivas e as arestas representam deslocamentos, incorporando restrições operacionais por meio de mascaramento de ações. A política de decisão é aprendida por meio do método *Advantage Actor-Critic* (A2C), utilizando uma função de recompensa que equilibra a distância percorrida, o transbordamento dos tanques, o volume coletado e o atendimento à demanda. Um estudo de caso em um cenário controlado foi conduzido para avaliar o comportamento do modelo. Os resultados indicam o potencial da abordagem para capturar padrões logísticos eficientes e adaptar-se a diferentes configurações operacionais, destacando sua aplicabilidade a problemas de maior complexidade e dinamismo.

Palavras-chave: *Graph Neural Networks*; Aprendizado por Reforço Profundo; Otimização Logística; Sistemas Dinâmicos e Descentralizados

1. MOTIVAÇÃO E CONTEXTUALIZAÇÃO

O atual cenário produtivo global demanda a conciliação entre lucratividade e sustentabilidade, o que incentiva o uso de matérias-primas de origem vegetal, animal e residual. No entanto, a principal limitação associada ao reaproveitamento dessas fontes, em substituição aos insumos fósseis, reside em sua distribuição geograficamente descentralizada. Nesse contexto, o Grupo de Intensificação, Modelagem, Simulação, Controle e Otimização de Processos (GIMSCOP) propõe o uso de *Redes de Unidades Modulares Autônomas/Automatizadas, Intensificadas e Sustentáveis* (RUMAS, ou RU+), que operam como biorrefinarias descentralizadas.

As RU+ consistem em redes de pequenas unidades totalmente automatizadas, operando de forma semiautônoma, próximas às fontes geradoras de biomassa e controladas por uma central responsável pelo monitoramento remoto da produção. Essas unidades vêm sendo desenvolvidas pelo GIMSCOP em duas frentes principais: (i) plataforma termoquímica, voltada ao processamento de resíduos agrícolas e urbanos por meio de pirólise rápida e torrefação; e (ii) plataforma fermentativa de amiláceos, com foco na produção de etanol e derivados a partir de batata-doce.

A operação de um sistema produtivo descentralizado impõe desafios adicionais à logística, especialmente quanto à coleta e ao escoamento dos produtos. Nesse contexto, novas variáveis devem ser consideradas em relação à logística tradicional, incluindo a capacidade produtiva e de armazenamento das unidades, a dinâmica de demanda na central, o tempo de prateleira dos produtos, bem como as restrições associadas ao carregamento e ao descarregamento de veículos.

Além disso, o caráter dinâmico do sistema logístico aumenta a complexidade do problema. Fatores como indisponibilidade de veículos, bloqueios rodoviários, congestionamentos, acidentes e condições climáticas adversas exigem que o sistema seja capaz de se adaptar continuamente. Dessa forma, torna-se fundamental o desenvolvimento de modelos capazes de recalcular, em tempo hábil, decisões ótimas ou quase ótimas diante de alterações no cenário operacional.

Tradicionalmente, o problema de roteamento veicular tem sido abordado por meio de métodos heurísticos ou determinísticos. Enquanto os métodos heurísticos buscam soluções aproximadas baseadas em regras

Realização:



inspiradas em fenômenos naturais, abordagens determinísticas, como a Programação Linear Inteira Mista (MILP), modelam matematicamente o problema visando à obtenção de soluções ótimas. Nesse contexto, Morais (2021), em seu trabalho no GIMSCOP, aplicou um MILP ao problema de roteamento veicular integrado à gestão de estoques. Embora resultados satisfatórios tenham sido obtidos em cenários de menor escala, observou-se a ocorrência de explosão combinatória à medida que o problema se tornava mais complexo, o que impossibilitava a obtenção de soluções ótimas em tempo de computação viável para cenários com múltiplos dias e dezenas de unidades.

A partir da análise da literatura e de estudos prévios, identificam-se três limitações principais das abordagens exatas para esse tipo de problema em larga escala: (i) rigidez e ausência de aprendizado, uma vez que soluções anteriores não contribuem para a resolução de novos cenários; (ii) baixa aplicabilidade em tempo real, devido ao elevado tempo computacional necessário para reotimização frente a pequenas mudanças no sistema; e (iii) limitada escalabilidade, decorrente do crescimento exponencial da complexidade com o aumento do número de variáveis.

Como alternativa, propõe-se o uso de *Graph Neural Networks* (GNNs) para a resolução do problema. Essas redes exploram a estrutura de grafos para propagação de informação, o que é particularmente adequado à modelagem do sistema em questão, em que unidades produtivas e centrais podem ser representadas como nós, e suas conexões como arestas contendo informações como distância e tempo. Adicionalmente, arquiteturas como *Graph Attention Networks* (GAT) permitem ponderar a relevância de cada nó com base em atributos operacionais, como o nível de estoque e a taxa de produção. A incorporação de técnicas de *Deep Reinforcement Learning* (DRL) possibilita, ainda, o treinamento do modelo a partir da realimentação do modelo com recompensas de acordo com o quão positivo foi seu comportamento com o ambiente, dispensando a necessidade de dados prévios.

Diante desse contexto, o presente trabalho tem como objetivo desenvolver e avaliar um modelo baseado em GNNs para o problema de roteamento veicular em sistemas descentralizados do tipo RU+. Após esta seção introdutória, o artigo está organizado da seguinte forma: inicialmente, apresenta-se a metodologia adotada, incluindo a descrição da arquitetura do modelo e do processo de treinamento; em seguida, são apresentados os resultados obtidos em um estudo de caso; por fim, discutem-se as conclusões e possíveis direções para trabalhos futuros.

2. METODOLOGIA

2.1. Fundamentação das Redes

O Problema de Roteamento Veicular (Vehicle Routing Problem – VRP) é uma classe de problemas de otimização combinatória, frequentemente modelados por Programação Linear Inteira Mista (MILP), cujo objetivo é determinar rotas ótimas para a visita de pontos geograficamente dispersos a partir de um depósito central, respeitando um conjunto de restrições operacionais e minimizando métricas como distância, custo ou tempo total de deslocamento (Laporte, 1992). Formalmente, o problema pode ser representado por um grafo, no qual os vértices correspondem aos pontos de atendimento e as arestas aos deslocamentos entre eles, devendo cada ponto ser visitado uma única vez e todas as rotas iniciarem e terminarem no depósito (Laporte, 1992). Variantes clássicas incluem a consideração de janelas de tempo (VRPTW), que impõem restrições temporais para atendimento de cada ponto, e de capacidade dos veículos (CVRP), que limitam a quantidade transportada em cada rota; a combinação dessas restrições dá origem ao Problema de Roteamento Veicular Capacitado com Janelas de Tempo (CVRPTW), cuja complexidade é classificada como NP-difícil (Kallehauge et al., 2005; Liu; Jiang, 2019).

Nesse contexto, a estrutura de grafo do problema motiva o uso de *Graph Neural Networks* (GNNs), que permitem a aprendizagem de representações vetoriais dos nós a partir da agregação das informações de suas vizinhanças. De forma geral, GNNs operam por meio de mecanismos de *message passing*, nos quais cada nó atualiza seu estado com base em suas próprias características e nas informações recebidas de seus nós adjacentes. No presente trabalho, considera-se uma variante dinâmica do problema, na qual múltiplas unidades produtivas (RU+) devem ser visitadas ao longo de um horizonte de planejamento discreto, em dias, respeitando as restrições de capacidade do veículo, o tempo máximo de operação e a dinâmica de produção contínua nas unidades.

2.2. Implementação do Modelo

O problema foi modelado em *Python*, utilizando a biblioteca *Pytorch*, como um grafo direcionado completo, no qual o nó de índice zero representa a central e os demais nós correspondem às unidades produtivas,

sendo as arestas responsáveis por representar os deslocamentos entre esses pontos. Adicionalmente, foi incluído um nó virtual para permitir que o modelo aprenda a decisão de encerramento do turno operacional.

As arestas do grafo são caracterizadas por atributos como distância euclidiana e tempo de deslocamento entre os nós, enquanto os vértices são descritos por um conjunto de *features* que representam o estado do sistema. Essas *features* incluem informações locais e globais, como volume armazenado, estimativa de produção futura, capacidade remanescente do veículo, demanda restante e tempo disponível no turno, conforme detalhado na Tabela 1.

Tabela 1. *Features* dos nós.

<i>Feature</i>	O que representa	Para que existe
É Central	Indica se o nó é a central [1] ou não [0]	Diferenciar funções dos nós
Volume Atual	Volume atual no tanque	Prioriza coleta em nós com maior volume
Localização Atual	Indica se o caminhão está [1] ou não está [0] neste nó	Permite decisões condicionais ao ponto atual
Volume Futuro	Volume que terei no tanque no início do próximo turno	Antecipação de transbordamento
Máximo Global	Maior volume que terei no início do próximo turno	Ajuda o modelo a identificar cenários críticos
Média Global	Média dos volumes que terei no início do próximo turno	Permite decisões mais equilibradas
Capacidade Restante	Capacidade restante do caminhão	Possibilita noção de restrição operacional
Demanda Faltante	Fração da demanda diária ainda não atendida	Faz o modelo balancear coleta e retorno à central
Tempo Restante	Tempo restante no turno	Permite decisões estratégicas (continuar, encerrar ou voltar)

No presente trabalho, optou-se pelo uso de duas camadas de convolução em grafos do tipo *TransformerConv*, que incorporam mecanismos de atenção para ponderar a importância relativa das conexões entre os nós. Essa abordagem permite que o modelo atribua maior relevância às unidades de maior criticidade operacional, como as com maior risco de transbordamento. As representações dos nós são propagadas por duas camadas ocultas de dimensão 64 e, posteriormente, utilizadas para gerar duas saídas distintas: uma associada à política (*policy head*) e outra à estimativa de valor (*value head*), conforme a estrutura de métodos *actor-critic*.

A saída da rede consiste em uma distribuição de probabilidade sobre os nós do grafo, que representa a probabilidade de seleção de cada ação possível. O espaço de ações do agente inclui a seleção do próximo nó a ser visitado, o retorno à central e o encerramento do turno. No entanto, nem todas as ações são viáveis em todos os estados. Para garantir a consistência operacional, foi implementado um mecanismo de mascaramento, no qual ações inviáveis recebem probabilidade nula. Entre as restrições consideradas, destacam-se: (i) proibição de revisitar nós já atendidos no mesmo turno/ciclo; (ii) restrições de tempo, que impedem deslocamentos que ultrapassem o limite de horário de um turno de trabalho; e (iii) regras operacionais, como a obrigatoriedade de retorno à central antes do encerramento do turno.

2.3. Estratégia de Treinamento

O modelo foi treinado por meio de aprendizado por reforço profundo (DRL), no qual um agente observa e interage com um ambiente ao longo do tempo, tomando decisões sequenciais e recebendo recompensas associadas à qualidade dessas decisões (Géron, 2025). O objetivo do agente é, portanto, aprender uma política que maximize a recompensa acumulada ao longo dos episódios.

A função de recompensa (*reward*), representada na Equação (1), foi projetada para incentivar a minimização da distância percorrida (*dist*), ao mesmo tempo em que recompensa a coleta eficiente de produto nas unidades (*carg*) e a descarga na central (*desc*). Além disso, penalizações são aplicadas em caso de transbordamento de tanques (*ntransb*) e de não atendimento à demanda de produtos na central (*deficit*). Os pesos associados à Equação (1) foram definidos de forma empírica, por meio de ajustes finos, buscando um comportamento de aprendizado inicial coerente com os objetivos do problema, sem a necessidade de um número elevado de repetições de treinamento.

$$reward = -0,05 \cdot deficit - 50 \cdot ntransb - 0,3 \cdot dist + 0,005 \cdot desc + 0,002 \cdot carg \quad (1)$$

O treinamento foi conduzido utilizando um método do tipo *Advantage Actor-Critic* (A2C), no qual a política (*actor*) e a função de valor (*critic*) são otimizadas simultaneamente. A atualização dos parâmetros é realizada com base no gradiente da política, ponderado pela vantagem estimada, definida como a diferença entre o retorno acumulado e o valor estimado pelo crítico (Géron, 2025).

Os episódios de treinamento consistem em simulações de múltiplos dias de operação. Em cada episódio, variáveis como a demanda diária, os volumes iniciais e a distribuição espacial das unidades são parcialmente aleatorizadas, o que aumenta a capacidade de generalização do modelo. As capacidades máximas do caminhão e dos tanques foram fixadas em 15.000 L e 10.000 L, respectivamente, e a taxa de produção de cada unidade foi definida em 41,67 L/h, considerando um turno operacional de 8 horas.

O treinamento foi conduzido utilizando o otimizador Adam, com taxa de aprendizado de 0,0005, fator de desconto de 0,99, indicando alta correlação entre as ações imediatas e futuras, e atualização dos parâmetros em *batches* de 20 episódios. Adicionalmente, foi incorporado um termo de entropia decrescente à função de perda, com o objetivo de incentivar a exploração nas etapas iniciais do treinamento.

3. RESULTADOS

3.1. ESTUDO DE CASO

A fim de avaliar o desempenho do modelo proposto, definiu-se um cenário controlado que permite analisar o comportamento do agente diante da distribuição espacial das unidades produtivas e das demandas associadas. O estudo de caso considera um sistema composto por uma central de coleta e cinco unidades produtivas, totalmente interconectadas e distribuídas em um plano bidimensional.

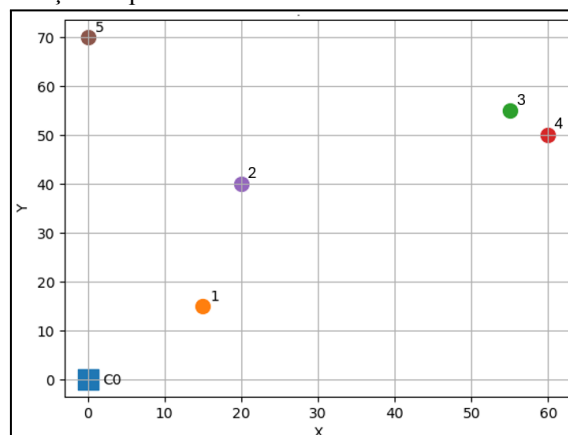
As coordenadas das unidades foram definidas de modo a criar um cenário com diferentes padrões logísticos: uma unidade localizada próxima à central, uma a uma distância intermediária e três unidades mais afastadas, sendo duas delas próximas entre si. Especificamente, as posições adotadas foram (0,0) para a central, (15,15) para a unidade próxima, (20,40) para a intermediária e (55,55), (60,50) e (0,70) para as unidades distantes, conforme ilustrado na Figura 1. Essa configuração permite avaliar se o modelo é capaz de identificar estratégias eficientes de agrupamento de visitas, priorizando rotas que minimizem deslocamentos desnecessários.

O horizonte de planejamento considerado foi de três dias, com demandas diárias fixadas em 4.000 L. As demais variáveis do sistema seguiram os parâmetros definidos na etapa de treinamento, incluindo capacidade máxima do caminhão de 15.000 L, capacidade dos tanques das unidades de 10.000 L, taxa de produção constante de 41,67 L/h e turno operacional de 8 horas.

Durante a execução do teste, o modelo foi avaliado em modo determinístico, selecionando, a cada etapa, a ação de maior probabilidade de acordo com a política aprendida. Essa abordagem permite observar diretamente a estratégia inferida pelo modelo, eliminando os efeitos estocásticos associados à amostragem de ações.

Ao longo da simulação, foram monitoradas variáveis relevantes para a análise do desempenho, incluindo o nó atual do agente, a carga transportada, o tempo decorrido, os volumes disponíveis em cada unidade e o atendimento à demanda diária. Dessa forma, torna-se possível avaliar qualitativamente as decisões tomadas pelo modelo, bem como sua capacidade de adaptação às condições operacionais impostas pelo cenário.

Figura 1. Posição no plano bidimensional das unidades no estudo de caso.



3.2. RESULTADOS E DISCUSSÕES

A fim de avaliar o desempenho do modelo proposto, foram realizadas múltiplas execuções do cenário descrito, o que permitiu analisar o comportamento do agente sob diferentes condições iniciais. A Tabela 2 apresenta exemplos representativos das rotas realizadas ao longo de três execuções distintas, evidenciando os caminhos percorridos em cada dia de operação e os volumes totais coletados.

Tabela 2. Concatenação dos resultados do estudo de caso.

Experimento	Dia	Caminho	Distância (km)	Coleta Total (L)
1	1	0 → 1 → 2 → 0 → FIM	91,43	13.518
	2	0 → 5 → 4 → 3 → 0 → FIM	218,10	15.000
	3	0 → 1 → 2 → 3 → 4 → 0 → FIM	169,96	10.795
2	1	0 → 1 → 2 → 0 → FIM	91,43	12.536
	2	0 → 5 → 3 → 4 → 0 → FIM	212,18	14.423
	3	0 → 1 → 2 → 0 → FIM	91,43	4.000
3	1	0 → 1 → 2 → 0 → FIM	91,43	11.690
	2	0 → 3 → 4 → 0 → FIM	162,96	15.000
	3	0 → 5 → 4 → 3 → 2 → 1 → 0 → FIM	225,10	14.649

A partir dos resultados obtidos, observa-se que o modelo apresenta um comportamento consistente de agrupamento espacial, priorizando a visita a unidades geograficamente próximas ao longo da mesma rota. Esse padrão evidencia a capacidade do modelo de capturar relações espaciais implícitas no problema, resultando em trajetórias mais eficientes do ponto de vista do deslocamento. Outro aspecto relevante refere-se à eficiência operacional das rotas, caracterizada pela realização de poucas viagens por dia, com uma única rota, em geral associada a altos volumes coletados por percurso, frequentemente próximos à capacidade máxima do veículo.

Adicionalmente, observa-se que o modelo tende a priorizar unidades mais próximas da central, deslocando-se para regiões mais distantes apenas quando necessário, seja para atender à demanda diária, seja para evitar transbordamento nos tanques das unidades. Esse comportamento indica a capacidade do agente de equilibrar múltiplos objetivos conflitantes, como a minimização da distância e a gestão de estoques distribuídos.

Apesar da eficiência observada na minimização de deslocamentos, essa estratégia pode não ser ideal em cenários em que a prioridade seja a coleta frequente de todas as unidades, independentemente da distância. Nesse contexto, a política aprendida pode favorecer a postergação da coleta em unidades mais distantes, o que pode impactar negativamente outros indicadores operacionais.

Por fim, verificou-se que o modelo apresenta desempenho mais consistente em cenários com maior número de unidades, sugerindo maior capacidade de generalização e de aproveitamento da estrutura de grafo em problemas de maior complexidade. Esse resultado reforça o potencial da abordagem proposta para aplicação em sistemas logísticos de maior escala, nos quais métodos tradicionais tendem a enfrentar limitações de desempenho e de adaptabilidade.

4. CONCLUSÕES

O presente trabalho propôs uma abordagem baseada em *Graph Neural Networks* (GNNs), combinada com aprendizado por reforço profundo, para a resolução do Problema de Roteamento Veicular em um contexto dinâmico e descentralizado, inspirada nas Redes de Unidades Modulares Autônomas/Automatizadas, Intensificadas e Sustentáveis (RU+). A modelagem do problema como um grafo, aliada ao uso de mecanismos de atenção e estratégias do tipo *actor-critic*, permitiu construir um modelo capaz de aprender políticas de decisão adaptativas a partir da interação com o ambiente.

A metodologia desenvolvida integrou, de forma consistente, a representação estrutural do problema, a incorporação de restrições operacionais por meio de mascaramento de ações e o uso de uma função de recompensa alinhada aos objetivos logísticos do sistema. Essa combinação permite ao modelo capturar relações espaciais e operacionais relevantes, ao mesmo tempo em que mantém a viabilidade das decisões ao longo da simulação.

Os resultados obtidos no estudo de caso indicam que a abordagem proposta é capaz de identificar estratégias logísticas coerentes com a estrutura do problema e apresenta comportamento adaptativo diante da distribuição espacial das unidades e das restrições impostas. Em especial, observa-se o potencial do modelo para explorar agrupamentos geográficos e otimizar o uso dos recursos disponíveis, características desejáveis em sistemas logísticos reais.

Apesar dos resultados promissores, o trabalho apresenta limitações inerentes à simplificação do ambiente considerado, especialmente no que diz respeito à escala do problema e à ausência de múltiplos agentes. Nesse sentido, diversas oportunidades de aprofundamento estão sendo exploradas em trabalhos futuros.

Como principal extensão, destaca-se a incorporação de múltiplos agentes, permitindo a modelagem de sistemas com mais de um veículo, o que representa um avanço significativo em direção a cenários reais de operação. Adicionalmente, sugere-se replicar os cenários desenvolvidos por Moraes (2021), com o objetivo de estabelecer uma comparação direta entre abordagens baseadas em otimização matemática clássica e em aprendizado de máquina, avaliando aspectos como a qualidade da solução e o tempo de computação.

Outras extensões relevantes incluem o escalonamento do modelo para um número maior de unidades, a introdução de perturbações no sistema – como bloqueios de vias e variações nas condições de transporte – e a incorporação de restrições adicionais, como o tempo de prateleira dos produtos, o que pode exigir políticas de coleta periódica. Tais avanços contribuiriam para tornar o modelo mais representativo de sistemas reais e para avaliar sua robustez em cenários mais complexos.

Agradecimentos: Ao CNPq, à CAPES e à FAPERGS pelos financiamentos ao longo de diversos projetos de pesquisa que permitem ao GIMSCOP desenvolver as RU+, bem como fomentar teses, dissertações e TCCs.

Referências

- A. Géron: Hands-On Machine Learning with Scikit-Learn and PyTorch: Concepts, Tools, and Techniques to Build Intelligent Systems (1ª Ed.), O'Reilly Media: 2025.
- B. Kallehauge et al.: Vehicle Routing Problem with Time Windows, Column Generation (G. Desaulniers, J. Desrosiers, and M. M. Solomon, Eds.), Springer US: 2005, p. 67–98.
- G. Laporte: The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms, European Journal of Operational Research v. 59(3), p. 345–358, 1992.
- R. Liu e Z. Jiang: A Hybrid Large-Neighborhood Search Algorithm for the Cumulative Capacitated Vehicle Routing Problem with Time-Window Constraints, Applied Soft Computing v. 80, p. 18–30, 2019.
- P. Moraes: Logística do Escoamento da Produção em Biorrefinarias Descentralizadas, Universidade Federal do Rio Grande do Sul, Porto Alegre: 2021.