

Aplicação do Aprendizado por Reforço para a otimização de parâmetros em Aprendizado Federado

Elisa Alves Veloso

12 de dezembro de 2025

Abstract

This work proposes combining Reinforcement Learning (RL) techniques with the Federated Learning (FL) paradigm, using RL to automatically update the training parameters employed in the federated process. In this context, the RL agent learns to adjust the number of participating clients based on performance metrics obtained throughout the training rounds, which serve as reinforcement signals. In this way, the goal is to integrate these two fields of study in order to optimize the federated training process and, consequently, improve the quality and robustness of the resulting global model.

Resumo

O presente trabalho propõe a combinação de técnicas de Aprendizado por Reforço (AR) com o paradigma de Aprendizado Federado (AF), utilizando AR para realizar a atualização automática dos parâmetros de treinamento empregados no processo federado. Nesse contexto, o agente de AR aprende a ajustar o número de clientes a partir das métricas de desempenho obtidas ao longo das rodadas, que funcionam como sinal de reforço. Dessa forma, busca-se integrar esses dois campos de estudo a fim de otimizar o processo de treinamento federado e, consequentemente, melhorar a qualidade e a robustez do modelo global gerado.

Palavras-Chave: Aprendizado por Reforço, Aprendizado Federado, Otimização de Hiperparâmetros, Seleção Adaptativa de Parâmetros, Modelos Distribuídos, Visão Computacional Agrícola, Detecção de Plantas Daninhas, Agricultura de Precisão, Deep Learning

1 Introdução

O Brasil é o maior produtor mundial de cana-de-açúcar. Sua produção corresponde a aproximadamente 37

A presença de plantas daninhas em canaviais é um processo natural. Contudo, devido à competição estabelecida entre as espécies, a produtividade da cultura pode ser reduzida em até 85

Tradicionalmente, a aplicação de defensivos agrícolas contra plantas daninhas é realizada por meio de grandes veículos pulverizadores, que distribuem herbicidas, generalistas ou específicos, sobre toda a plantaçāo. Essa prática pode ocasionar diversos prejuízos. Primeiramente, parte do defensivo pode atingir a cultura principal, comprometendo seu crescimento, reduzindo a qualidade da produção e gerando preocupações ao consumidor quanto à segurança do alimento. Além disso, o escoamento dos pesticidas pelo solo pode contaminar lençóis freáticos, acarretando impactos ambientais significativos. Soma-se a isso o risco de seleção de plantas daninhas resistentes, o que exige, posteriormente, o uso de produtos mais potentes e intensifica o ciclo de resistência [AN19].

Dessa forma, a identificação precisa de plantas daninhas pode trazer benefícios expressivos aos produtores, permitindo que a aplicação de defensivos seja direcionada apenas às áreas efetivamente infestadas. Assim, reduz-se o desperdício de insumos, minimizam-se impactos ambientais e aprimora-se o processo de tomada de decisão [AN19].

Atualmente, as soluções inteligentes voltadas à identificação de plantas daninhas operam, em grande parte, com modelos centralizados, treinados a partir de dados fornecidos por grandes usinas. Esse formato implica a exposição de informações sensíveis dos produtores, uma vez que os dados de colheita são enviados para compor o modelo global. Como alternativa, destaca-se a aprendizagem federada, que propõe o treinamento de modelos diretamente nos dispositivos dos produtores, compartilhando apenas os pesos atualizados com um servidor central. Essa abordagem permite construir um modelo eficiente, ao mesmo tempo em que preserva a privacidade e a segurança dos dados sensíveis.

O objetivo geral deste trabalho é investigar e propor uma abordagem que integra Aprendizado Federado (FL) com Aprendizado por Reforço (AR), de forma que: (a) o agente aprenda automaticamente o melhor conjunto de clientes para conduzir os treinamentos federados; (b) as ações representem ajustes nesses conjuntos de clientes ao longo das rodadas do Aprendizado Federado; (c) o sinal de reforço seja uma métrica de desempenho (acurácia do modelo global), juntamente com a eficiencia com esse conjunto selecionado; (d) a aplicação alvo seja a agricultura de precisão, especialmente o problema de identificar plantas daninhas em canaviais usando visão computacional; (e) O trabalho aprofunde o uso do AR como ferramenta para a otimização automática de FL, destacando-o como uma nova técnica e como uma nova aplicação.

Nos últimos anos, o crescimento exponencial de dispositivos conectados impulsionou a geração de grandes volumes de dados distribuídos entre diferentes fontes e regiões. Nesse contexto, o Aprendizado Federado (Federated Learning – FL) emergiu como uma alternativa promissora para o treinamento colaborativo de modelos de aprendizado de máquina sem a necessidade de centralização dos dados, contribuindo para a preservação da privacidade e para a redução de custos de transmissão. Contudo, apesar do seu potencial, o desempenho do FL depende diretamente de um conjunto de hiperparâmetros que governam o processo de treinamento, tais como taxa de aprendizado, quantidade de clientes participantes por rodada e número de épocas locais. Ajustar tais parâmetros manualmente é um processo custoso, pouco escalável e sensível à heterogeneidade dos dispositivos e dos dados envolvidos [WZW22], [SAA24], [XZL24], [JWM23].

Nesse cenário, o Aprendizado por Reforço (AR) surge como uma abordagem capaz de automatizar a tomada de decisões sequenciais, permitindo que um agente aprenda políticas de otimização baseadas em interações com o ambiente. Trabalhos recentes têm demonstrado que integrar AR ao FL pode trazer ganhos significativos de desempenho, principalmente em tarefas de seleção de clientes e ajuste dinâmico de parâmetros ([WZW22]; [CJC24]). Tais pesquisas mostram que o uso de Q-learning, DDQN e métodos baseados em políticas contínuas, tais como PPO, permitem reduzir o tempo de convergência, melhorar a acurácia global e lidar com a heterogeneidade dos dados distribuídos [SAA24], [XZL24].

Apesar dos avanços, a aplicação conjunta de FL e AR ainda é pouco explorada em domínios

agrícolas, especialmente em cenários relacionados à agricultura de precisão, à identificação automática de plantas daninhas, e à proteção dos dados sensíveis de grandes produtores. Com base nessa lacuna, o presente trabalho propõe um modelo no qual um agente de AR atua no servidor agregador, ajustando automaticamente os clientes adicionado ao treinamento do aprendizado federado em tempo de execução. A proposta busca utilizar RL para otimizar o processo federado como um todo, visando maior estabilidade, eficiência e qualidade do modelo resultante.

A integração de AR ao FL para visão computacional aplicada a canaviais visa ainda fortalecer a robustez do modelo global e ampliar sua capacidade de generalização entre diferentes produtores rurais, mantendo a privacidade dos dados de imagem e reduzindo o custo de configuração manual do sistema. Assim, espera-se contribuir com o avanço científico sobre automação inteligente de Aprendizado Federado, além de oferecer uma solução prática e escalável para desafios reais de classificação agrícola.

2 Fundamentação Teórica

A literatura recente em Federated Learning tem destacado que, embora o paradigma possibilite o treinamento colaborativo de modelos sem centralizar dados, sua eficácia reduz-se significativamente em ambientes compostos por dispositivos heterogêneos, especialmente no contexto de AIoT. Em [CJC24], os autores investigam a problemática da queda de desempenho decorrente de limitações de hardware, variações de memória, capacidade de processamento e incertezas operacionais. A heterogeneidade estrutural e operacional restringe o uso de modelos globais homogêneos, que, ao serem distribuídos para dispositivos incapazes de suportá-los, levam a falhas de treinamento, degradação de acurácia e desperdício de comunicação. Nessa linha, os autores estruturam uma solução que busca conciliar eficiência computacional, respeito a restrições de recursos e preservação de desempenho global. Para isso, o artigo introduz o AdaptiveFL, um framework que combina poda de modelo em granularidade fina e uma estratégia de seleção de dispositivos baseada em aprendizagem por reforço.

A problemática central discutida pelos autores em [WZW22] refere-se à forte limitação de recursos em dispositivos de borda, que apresentam capacidades computacionais desiguais, níveis variados de conectividade e comportamentos instáveis ao longo do tempo. Em ambientes reais, muitos desses dispositivos não conseguem executar localmente um modelo dentro dos prazos exigidos, o que gera atrasos significativos, falhas de resposta e perda de sincronização global. Esse cenário compromete a velocidade de convergência e reduz a qualidade do modelo final. Os autores observam que estratégias tradicionais, como seleção aleatória ou seleção baseada apenas na capacidade de processamento, são insuficientes porque ignoram fatores dinâmicos, como flutuações de carga, energia residual e instabilidade de rede. Em resposta a esse desafio, o artigo estrutura uma solução orientada por aprendizagem profunda por reforço, com o objetivo de otimizar a escolha dos dispositivos a cada rodada de FL. A proposta insere a seleção de clientes em um processo decisório sequencial, no qual o sistema aprende políticas capazes de identificar combinações de dispositivos que melhorem a eficiência sem comprometer a diversidade de dados.

A pesquisa em Federated Learning tem evidenciado que a heterogeneidade entre dispositivos participantes representa um dos maiores obstáculos à eficiência e à estabilidade do treinamento distribuído. O artigo [SAA24] aborda precisamente essa problemática, destacando que as diferenças entre capacidades computacionais, velocidades de comunicação, tamanhos de dados e padrões de energia comprometem a convergência global do modelo. Em ambientes heterogêneos, alguns clientes tornam-se gargalos por apresentarem lentidão excessiva ou falhas de conexão, enquanto outros podem contribuir de forma desproporcional, gerando desequilíbrios no processo de agregação. Abordagens tradicionais que selecionam clientes de forma aleatória ou com base em parâmetros fixos não são

capazes de lidar adequadamente com a dinamicidade do ambiente, resultando em desperdício de recursos e degradação da acurácia final do modelo. Com o objetivo de mitigar esses desafios, os autores estruturam uma solução adaptativa que utiliza aprendizagem por reforço profundo para otimizar a seleção de clientes em cada rodada de treinamento federado

[GYH⁺22] também aborda a forte influência da heterogeneidade dos dados e a consequente necessidade de ajustar cuidadosamente os hiperparâmetros de treinamento. A literatura revisada evidencia que algoritmos clássicos de FL, como FedAvg, apresentam comportamento instável sob distribuições não independentes e não idênticas, ocasionando divergência entre modelos locais, lentidão na convergência e degradação do desempenho global. A problemática também envolve características estruturais próprias do FL. Por um lado, a colaboração entre instituições médicas, embora essencial para superar a limitação de dados, é restringida por regulamentações de privacidade que proíbem a centralização das informações. Por outro, a diversidade de instituições leva a distribuições de dados profundamente distintas, de modo que o ajuste manual e estático de hiperparâmetros dificilmente se adapta simultaneamente a todos os contextos locais. Essa combinação de restrições faz com que pequenas escolhas, como taxas de aprendizado, número de iterações locais e pesos de agregação, afetem diretamente a estabilidade do processo federado. Para enfrentar esses desafios, o artigo propõe o Auto-FedRL, um método inovador que formula a otimização de hiperparâmetros como um problema de decisão sequencial resolvido por aprendizagem por reforço. A solução é estruturada de modo que um agente de RL opere conjuntamente com o ciclo do FL, ajustando os hiperparâmetros a cada rodada com base na redução relativa da perda de validação observada nos clientes.

3 Metodologia

A metodologia adotada neste trabalho foi estruturada em etapas sequenciais que abrangem desde a seleção dos dados até a análise final do modelo treinado por Aprendizagem Federada. Cada fase foi planejada para garantir a consistência experimental e a análise adequada dos resultados. As etapas seguidas são descritas a seguir.

Inicialmente, foram selecionados os conjuntos de dados adequados ao problema de identificação de plantas daninhas em lavouras de cana-de-açúcar. A escolha considerou critérios como disponibilidade pública em [Dom25], diversidade de classes, qualidade das imagens e representatividade das condições reais de campo. O dataset é composto por duas classes de imagens, totalizando 1.046 amostras, sendo elas "daninha" e "nao-daninha". Elas têm formato .jpeg e têm dimensões 240x320.

Após a seleção, os conjuntos de dados passaram por etapas de pré-processamento para garantir consistência e padronização. A fim de preparar o material de forma adequada para os experimentos centralizados e federados, o processo incluiu:

1. Normalização e redimensionamento das imagens;
2. Remoção de amostras corrompidas ou redundantes;
3. Aplicação de técnicas de aumento de dados;
4. Divisão dos dados em subconjuntos de treino, validação e teste.

Na etapa seguinte, o modelo foi treinado utilizando o paradigma de Aprendizagem Federada. O dataset foi particionado entre clientes simulados, que executaram treinamentos locais independentes. Após cada rodada de treinamento, os gradientes ou parâmetros atualizados foram enviados ao servidor agregado central, onde um algoritmo de agregação consolidou as contribuições individuais para formar um modelo global atualizado.

O modelo federado com um agente de Aprendizado por Reforço foi implementado, com o objetivo de ajustar hiperparâmetros específicos desse contexto, como o número de rodadas, taxa de aprendizado local, número de épocas por cliente e proporção de clientes participantes por rodada. Essa fase buscou maximizar o desempenho federado respeitando as restrições de privacidade e a heterogeneidade entre os dados locais.

O simulador para a Aprendizagem Federada é framework Flower, uma biblioteca que oferece recursos para executar experimentos de FL em larga escala e considerar cenários de dispositivos de FL ricamente heterogêneos. Essa ferramenta abstrai as complexidades de comunicação e orquestração do aprendizado federado, de modo que a implementação deste trabalho tenha sido focada apenas na integração da estratégia de RL ao servidor Federado.

O problema de orquestração de clientes foi modelado como um Processo de Decisão de Markov (MDP). O Estado é definido por um vetor contendo o índice da rodada atual e a acurácia global da rodada anterior. A Ação corresponde à seleção discreta do número de clientes participantes ($K \in [1, 5]$). A Recompensa é formulada como a acurácia obtida na rodada atual penalizada linearmente pelo número de clientes utilizados $r = Acc - 0.01 * K$, sendo K o número de clientes selecionados para treinar, incentivando a eficiência de recursos. O agente utiliza o algoritmo Deep Q-Network (DQN) para aproximar a função de valor $Q(s,a)$ e aprender a política ótima de seleção de clientes.

O modelo de classificação compartilhado é uma CNN simples. Em cada rodada de comunicação, um agente DQN, residindo no servidor, observa o estado atual do sistema, isto é, rodada e acurácia prévia, e determina o subconjunto ideal de clientes a serem convocados.

```
state = [round_num, last_accuracy]
action_idx = rl_agent.act(state)
num_clients_to_sample = action_idx + 1
```

A função de valor $Q(s,a)$ foi aproximada por uma Rede Neural Perceptron Multicamadas (MLP) com duas camadas ocultas de 24 neurônios cada e ativação ReLU, recebendo o vetor de estado como entrada e retornando os Q-values para cada ação possível.

```
class DQN(nn.Module):
    def __init__(self, state_dim, action_dim):
        self.fc1 = nn.Linear(state_dim, 24)
        self.fc2 = nn.Linear(24, 24)
        self.fc3 = nn.Linear(24, action_dim)
```

Utilizou-se uma política ϵ -greedy para a seleção de ações. O parâmetro ϵ iniciou em 1.0, decaindo a uma taxa de 0.995 por episódio até um mínimo de 0.01, garantindo uma transição gradual de exploração aleatória para exploração da política aprendida.

```
if np.random.rand() <= self.epsilon:
    return random.randrange(self.action_dim)
return torch.argmax(act_values[0]).item()
```

O objetivo do agente é maximizar uma função de recompensa que balança a acurácia do modelo global com o custo computacional de acionar múltiplos clientes. Assim sendo, tem-se que

```
reward = current_accuracy - (0.01 * num_clients_to_sample)
```

Dessa maneira, o agente recebe uma recompensa maior se a acurácia for alta, mas a adição de clientes reduz o valor final da recompensa, de modo a representar a penalidade pelo custo. Assim sendo, e ele usar muitos clientes e a acurácia subir pouco, a recompensa diminui. Se ele usar poucos

clientes e a acurácia cair, a recompensa diminui. Por fim, o ponto ideal é um valor que maximize a acurácia mantendo o mínimo número de clientes possível.

O treinamento ocorre em um loop onde os clientes treinam localmente, o servidor agrupa os parâmetros via FedAvg, e o agente RL atualiza sua política baseada na recompensa recebida.

```
self.memory.append((state, action, reward, next_state, done))
minibatch = random.sample(self.memory, batch_size)
for state, action, reward, next_state, done in minibatch:
    target = reward
    if not done:
        next_state_tensor =
            torch.FloatTensor(next_state)
        target = (reward + self.gamma * torch.max(self.model
            (next_state_tensor)[0]))
```

Concluídos os treinamentos, os resultados obtidos pela abordagem federada com aprendizado por reforço foram analisados e comparados, baseados em 3 experimentos: o primeiro com 10 rodadas de treinamento, o segundo com 25 rodadas e o último com 100.

Por fim, foram discutidas as principais conclusões do estudo, destacando a eficácia da Aprendizagem Federada para tarefas envolvendo dados sensíveis no contexto agrícola. Também foram apresentados direcionamentos para trabalhos futuros, como aplicação da metodologia a datasets maiores, inclusão de arquiteturas mais avançadas, integração com sensores embarcados e testes em ambientes reais de produção.

4 Resultados

Nos experimentos, o sistema demonstrou aprendizado consistente do modelo global, um aumento consistente na acurácia. À mesma medida, as ações do agente nos três casos resultou no aumento da recompensa recebida, que aumentou à medida que o número de rodadas também cresceu.

Do ponto de vista do comportamento do agente, os três resultados demonstram muita oscilação a cada rodada, o que caracteriza o fato de o agente ainda estar agindo muito exploratória, e não utilizando a memória desenvolvida a cada rodada. Isso pode indicar que o valor epsilon está muito alto, e por tanto prioriza a exploração das ações.

A análise das recompensas sugere que políticas de seleção esparsas, selecionar menos clientes, podem ser ótimas para este cenário, dado que o ganho marginal de acurácia ao adicionar todos os clientes não superou o custo computacional imposto pela função de recompensa

Mesmo explorando aleatoriamente, pode-se ver quais situações geraram melhores recompensas, o que indica que o agente aprenderá melhor em futuras rodadas.

Rodadas	Acurácia	Recompensa	Custo
5 e 6 (5 clientes)	73-76%	0.68 - 0.71	Alto (5 clientes = -0.05)
8 (1 cliente)	81%	0.80	Baixo (1 cliente = -0.01)

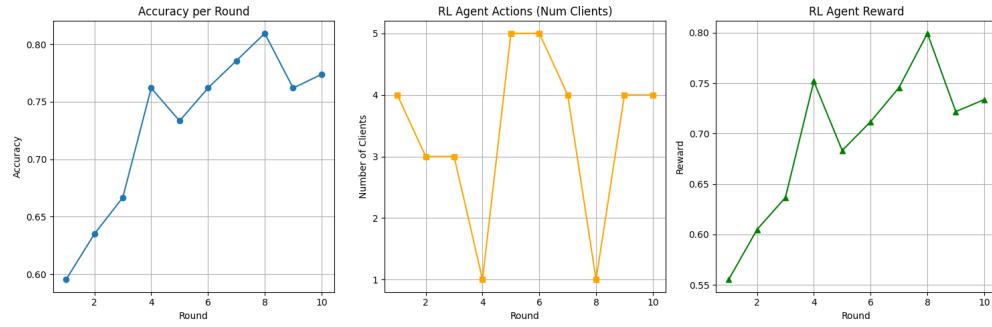


Figura 1: Experimento rodado em 10 rodadas



Figura 2: Experimento rodado em 25 rodadas



Figura 3: Experimento rodado em 100 rodadas

5 Conclusão

[h!] Baseado nos experimentos presentes nesse trabalho, pôde-se verificar que o Aprendizado por Reforço pode ser um grande aliado na seleção de parâmetros no processo de Aprendizagem Federada.

A integração de Aprendizado por Reforço (RL) no processo de orquestração do Aprendizado Federado demonstrou ser uma abordagem viável e promissora para a otimização dinâmica de recursos. Os experimentos realizados com o agente DQN evidenciam que a seleção de clientes não precisa ser estática ou aleatória, mas pode ser adaptada em tempo real com base no estado atual do modelo global.

O agente foi capaz de identificar que, para o cenário e dataset testados, a utilização massiva de clientes nem sempre traduzia-se em ganhos proporcionais de acurácia que justificassem o custo computacional. As maiores recompensas foram obtidas em rodadas com seleção esparsa de clientes, sugerindo que o RL pode reduzir significativamente a sobrecarga de comunicação sem degradar a performance do modelo.

A modelagem via MDP permite que o sistema aprenda políticas complexas que podem variar conforme o estágio do treinamento, embora experimentos de longa duração sejam necessários para observar a convergência total dessa política.

Referências

- [AN19] F. C. R. Amaral and M. P. Nepomuceno. Períodos de interferência de plantas daninhas na cana-de-açúcar cultivada em sistema de mudas pré-brotadas. *Planta Daninha*, 37, 2019.
- [CJC24] Z. Chen Y. Yang X. Xie Y. Liu C. Jia, M. Hu and M. Chen. Adaptivefl: Adaptive heterogeneous federated learning for resource-constrained aiot systems. *Proceedings of the 61st ACM/IEEE Design Automation Conference (DAC)*, page 1–6, 2024.
- [Dom25] Dhirendra Kumar Domah. Sugarcane and weed dataset. Kaggle Dataset, 2025.
- [GYH⁺22] Pengfei Guo, Dong Yang, Ali Hatamizadeh, An Xu, Ziyue Xu, Wenqi Li, Can Zhao, Daguang Xu, Stephanie Harmon, Evrim Turkbey, Baris Turkbey, Bradford Wood, Francesca Patella, Elvira Stellato, Gianpaolo Carrafiello, Vishal M. Patel, and Holger R. Roth. Auto-fedrl: Federated hyperparameter optimization for multi-institutional medical image segmentation, 2022.
- [JWM23] S. Cui L. Che L. Lyu D. Xu J. Wang, X. Yang and F. Ma. Towards personalized federated learning via heterogeneous model reassembly. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.
- [SAA24] H. O. Alanazi S. Ahmed, M. AbdelBaky and M. Amoon. Adaptive federated learning with reinforcement learning-based client selection for heterogeneous environments. *Computers and Electrical Engineering*, 117, 2024.
- [WZW22] Z. Liu B. Ding Y. Sun W. Zhang, Y. Jiang and Y. Wang. Adaptive client selection in resource-constrained federated learning systems: A deep reinforcement learning approach. *Computer Communications*, 188:56 – 66, 2022.
- [XZL24] L. Wang X. Zhao and J. Liu. Fedppo: Reinforcement learning-based client selection for federated learning with heterogeneous data. *preprint*, 2024.

6 Observações

O ChatGPT foi utilizado neste trabalho para auxiliar na tradução de conteúdos e na geração de referências em formato .bib.