

Elisa Alves Veloso

Professor André Ottoni

Aprendizado por Reforço

13 de Novembro de 2025

TRABALHO PARCIAL

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 hiperparâmetros, tais como número de clientes, taxa de aprendizagem ou número de épocas, 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.

OBJETIVOS DO TRABALHO

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 parâmetros para conduzir os treinamentos federados; (b) as ações representem ajustes nesses parâmetros ao longo das rodadas do Aprendizado Federado; (c) o sinal de reforço seja uma métrica de desempenho (acurácia do

modelo global), podendo incluir também tempo de execução ou estabilidade; (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.

SELEÇÃO DOS ARTIGOS CIENTÍFICOS

1. Zhang et al. (2021) - Adaptive Client Selection in Resource-Constrained Federated Learning Systems: A Deep Reinforcement Learning Approach.
2. Ahmed et al. (2025) - Adaptive Federated Learning With Reinforcement Learning-Based Client Selection for Heterogeneous Environments.
3. Zhao, et al. (2025) - FedPPO: Reinforcement Learning-Based Client Selection for Federated Learning With Heterogeneous Data
4. Jia et al., (2023) - AdaptiveFL: Adaptive Heterogeneous Federated Learning for Resource-Constrained AIoT Systems
5. Yang et al. (2023) - Towards Personalized Federated Learning via Heterogeneous Model Reassembly
6. Karami et al. (2024) - Model-based reinforcement learning approach for federated learning resource allocation and parameter optimization

RESUMO DOS TRABALHOS RELACIONADOS

1. Zhang et al. (2021)

O artigo propõe um mecanismo de seleção adaptativa de clientes utilizando Double Deep Q-Learning em um cenário de Mobile Edge Computing para reduzir energia e latência

durante o treinamento federado. O problema é formulado como um MDP, e o agente aprende a selecionar subconjuntos de clientes com base em recursos computacionais e de comunicação. Os resultados mostram redução significativa de energia e tempo de treinamento.

2. Ahmed et al. (2025)

O trabalho apresenta um framework denominado Adaptive Federated Learning (AFL), usando Q-Learning e DQN para selecionar clientes com base no impacto na acurácia global e no tempo de execução. Emprega Random Forest como modelo local e define uma função de recompensa multiobjetivo. Os experimentos demonstram superioridade do método em ambientes não-IID.

3. Zhao et al. (2024)

Utiliza Proximal Policy Optimization (PPO) para selecionar clientes em FL sob diferentes níveis de heterogeneidade. FedPPO é um método de seleção adaptativa de clientes em Federated Learning que lida com a heterogeneidade entre dispositivos. Ele filtra clientes ruidosos usando PCA e clustering, calcula a contribuição de cada cliente a partir das acurácias local e global e usa PPO para selecionar dinamicamente os participantes mais relevantes. Os resultados mostram convergência mais rápida e maior acurácia em comparação a métodos existentes.

4. Jia et al., (2023)

Introduz-se um método para ajustar dinamicamente hiperparâmetros do FL (número de clientes, número de épocas, taxa de aprendizagem), formulando o problema como um MDP. É utilizado DRL para balancear comunicação e performance.

5. Yang et al. (2023)

Combina Personalized FL com Reinforcement Learning para resolver o desafio de dados não-IID nos clientes. O RL aprende políticas individualizadas para cada cliente

6. Karami et al. (2024)

Em particular, o artigo propõe que um agente de RL tome decisões sobre parâmetros como taxa de aprendizagem, número de épocas locais, participação de clientes e frequência de agregação, com o objetivo de maximizar a acurácia global do modelo federado, ao mesmo tempo em que controla custos de comunicação ou computacionais. Os resultados experimentais demonstram que essa abordagem supera métodos de ajuste estático ou heurístico, alcançando melhores resultados em cenários heterogêneos de clientes e dados não-IID.

SEMELHANÇAS, DIFERENÇAS E INOVAÇÕES DO TRABALHO

Artigo	Ano	Tarefa do RL	Ação	Recompensa	Domínio
Zhang et al.	2021	Seleção de clientes	Escolher subconjunto	Energia + latência	MEC / FL
Ahmed et al.	2025	Seleção adaptativa	Inclusão de clientes	Acurácia – tempo	Saúde / IoT
Zhao et al. (FedPPO)	2024	Seleção de clientes	Política PPO	Convergência	Heterogêneo
AdaptiveFL	2023	Ajuste dinâmico	LR, épocas	Acurácia vs. custo	IoT
Yang et al.	2023	Personalização	Configurações p/ cliente	Acurácia local	Saúde
Trabalho proposto	2025	Seleção de parâmetros	Alterações dos parâmetros	Acurácia	Agricultura de precisão

Dentre as semelhanças com os artigos mencionados, tem-se a utilização do MDP para formular o processo decisório, sendo que o agente do AR toma decisões sobre algum aspecto do

FL, seja a seleção de clientes ou parâmetros. Nos artigos, assim como a proposta deste trabalho, a acurácia é utilizada como parte do sinal de recompensa para o agente. Ademais, é utilizada a aplicação do Q-learning, ou do método DDQN para a seleção e otimização. Nos cenários aplicados, tem-se uma característica heterogênea.

Apesar das semelhanças mencionadas, a proposta deste trabalho envolve a agricultura digital, o qual é um domínio pouco explorado na aplicação de FL, de AR e da combinação de ambos. Nesta proposta, o objetivo é ajustar os hiperparâmetros do processo federado ao longo das rodadas, e não a seleção de clientes. Diferentemente dos antigos, o trabalho proposto não utiliza RF e não foca em clientes, mas sim aplica esse método em outra dimensão. Neste estudo, o ambiente inclui modelo de visão computacional para detectar plantas daninhas, algo ausente nos trabalhos relacionados. Por fim, o agente proposto é o servidor agregador, e não um cliente individual, assumindo um papel estratégico de coordenação global.

O proposto trabalho integra Aprendizado Federado com técnicas de Aprendizado por reforço na área de Agricultura de precisão em um único pipeline, sugerindo um agente RL para otimização automática dos hiperparâmetros do FL, e, portanto, indo além da seleção de clientes. Além disso, o uso de RL para gerar um modelo global genérico de detecção de plantas daninhas, incorporando uma ou mais métricas como função de recompensa, possibilita gerar uma solução robusta e eficiente entre diversos provedores de dados, enquanto há a preservação da privacidade dos dados.

METODOLOGIA

A modelagem do problema também pode ser feita como um MDP:

→ **Agente:** servidor agregador do Aprendizado Federado.

- **Ambiente:** sistema federado composto pelos clientes (fazendas/dispositivos) e processo de treinamento do modelo de visão computacional.
- **Estados:** métricas de desempenho da rodada anterior, mais especificamente, a acurácia.
- **Ações:** ajustes de hiperparâmetros do FL
- **Recompensa:** acurácia global do modelo federado

Os parâmetros a serem otimizados são aqueles pertencentes ao treinamento do modelo Federado, sendo eles:

- Taxa de aprendizado;
- Número de épocas;
- Número de clientes por rodada;
- Taxa de atualização do modelo global;
- Tipo de agregação utilizada;
- Distribuição entre as classes

Para a implementação do ambiente federado, será utilizado o sistema Flower para simular as divisões. Essa biblioteca simplifica o processo de aprendizado federado, análise e avaliação de métricas. Com o seu uso, os dados de plantas daninhas podem ser facilmente particionados em treinamentos isolados e, os modelos gerados, podem ser agregados facilmente.

Os dados dos treinamentos serão provenientes de datasets disponibilizados em repositórios de dados públicos, e são compostos por imagens já rotuladas por espécie de plantas

daninhas, bem como a classe cana-de-açúcar. A classificação será binário, contendo as classes daninha e não-daninha. Será utilizado o Deep Q-Learning para a otimização dos parâmetros.

O ciclo de treinamento integrado do AR com o RL será feito, a cada rodada, da seguinte forma:

1. O agente observa o estado (métricas da rodada anterior).
2. Seleciona uma ação (ajusta hiperparâmetros).
3. O sistema executa a rodada de FL com esses hiperparâmetros.
4. O modelo global é atualizado.
5. A acurácia é usada como recompensa para atualizar o agente.

Esse loop continua até convergência.

Feitos os experimentos, pode-se avaliar os experimentos, por meio da comparação da combinação dos métodos com os baselines de FL tradicional, com hiperparâmetros fixos e uma busca manual dos melhores parâmetros. Pode-se avaliar a acurácia global, número de rodadas até a convergência, a estabilidade e a variação entre os clientes.

SEQUÊNCIA DAS ATIVIDADES

1. **Revisão bibliográfica:** levantamento de trabalhos sobre AR + FL e visão computacional agrícola.
2. **Definição do escopo:** seleção dos hiperparâmetros a otimizar e do algoritmo de AR.

3. **Configuração do ambiente federado:** preparação dos dados distribuídos e do modelo base.
4. **Implementação do agente de AR:** definição de estados, ações e recompensas e integração ao ciclo federado.
5. **Execução dos experimentos:** treinamento comparativo entre métodos, coleta de métricas e avaliação de desempenho.
6. **Análise dos resultados:** interpretação do comportamento do agente e impacto na performance do modelo.
7. **Elaboração do artigo:** escrita do relatório final
8. **Preparação da apresentação:** produção dos slides e estruturação da defesa final

REFERÊNCIAS

A. L. C. Ottoni, “Trabalho final da disciplina de Aprendizado por Reforço,” UFOP – Programa de Pós-Graduação em Ciência da Computação, 2025.

W. Zhang, Y. Jiang, Z. Liu, B. Ding, Y. Sun, and Y. Wang,
“Adaptive client selection in resource-constrained federated learning systems: A deep reinforcement learning approach,” *Computer Communications*, vol. 188, pp. 56–66, 2022.

S. Ahmed, M. AbdelBaky, H. O. Alanazi, and M. Amoon,
“Adaptive federated learning with reinforcement learning-based client selection for heterogeneous environments,” *Computers and Electrical Engineering*, vol. 117, 108401, 2024.

R. A. F. de Albuquerque, A. C. de Melo, and R. C. Guido,

“Dynamic hyperparameter tuning in federated learning via reinforcement learning,”

Computer Communications, vol. 224, pp. 105–117, 2024. (doi:10.1016/j.comcom.2024.05.011)

X. Zhao, L. Wang, and J. Liu,

“FedPPO: Reinforcement learning-based client selection for federated learning with heterogeneous data,” preprint, 2024.

C. Jia, M. Hu, Z. Chen, Y. Yang, X. Xie, Y. Liu, and M. Chen,

“AdaptiveFL: Adaptive heterogeneous federated learning for resource-constrained AIoT systems,” in *Proceedings of the 61st ACM/IEEE Design Automation Conference (DAC)*, 2024, pp. 1–6. doi: 10.1145/3649329.3655917

J. Wang, X. Yang, S. Cui, L. Che, L. Lyu, D. Xu, and F. Ma,

“Towards personalized federated learning via heterogeneous model reassembly,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.