Universidade de Pernambuco
Programa de Pós-Graduação em Engenharia da Computação (PPGEC)

Proposta de Tese de Doutorado

Área: Inteligência Computacional

Título: Uma Abordagem Baseada em Aprendizado Federado baseada em Clonagem de Modelos

Orientador – Bruno José Torres Fernandes (bjtf@ecomp.poli.br)

Descrição – O avanço da tecnologia e a coleta massiva de dados têm impulsionado o desenvolvimento de algoritmos de aprendizado de máquina cada vez mais complexos e eficientes [1]. No entanto, muitas vezes esses dados são distribuídos em diferentes locais, o que torna difícil a tarefa de treinar um modelo com essas informações [2]. Além disso, o compartilhamento desses dados pode violar a privacidade das pessoas ou empresas envolvidas [3].

O aprendizado federado é um paradigma recente [4] que permite o treinamento de modelos de aprendizado de máquina em conjuntos de dados distribuídos e descentralizados, sem a necessidade de compartilhar os dados diretamente entre os participantes. O aprendizado federado permite então que múltiplas entidades (por exemplo, dispositivos móveis, sensores, servidores em diferentes locais) colaborem para treinar um modelo sem compartilhar seus dados brutos. Esta abordagem promove a colaboração entre diferentes entidades, ao mesmo tempo em que protege a privacidade dos dados e minimiza os custos associados à comunicação e ao armazenamento centralizado e pode ser utilizada em diversos cenários, desde segurança urbana até cidades inteligentes [5].

Naturalmente que o aprendizado federado também apresenta alguns desafios. No caso da eficiência, por exemplo, embora possa haver um aprendizado prévio distribuído reduzindo o custo na parte central da arquitetura, a comunicação pode se tornar um gargalo considerando o número de dispositivos conectados [6]. O aspecto da colaboração também pode se tornar um ponto a ser trabalhado quando não existe um equilíbrio entre as partes na produção do conhecimento [7]. Além disso, problemas como tolerância a falhas, para garantir que cada dispositivo participante seja contabilizado e que o desempenho não seja comprometido, e diversidade de dados precisam ser tratados para um melhor desempenho das técnicas de aprendizado federado.

Este projeto propõe o desenvolvimento de uma solução para lidar com a heterogeneidade dos dados e a distribuição não uniforme dos conjuntos de dados entre os participantes, buscando melhorar a qualidade e a generalização do modelo treinado nas áreas mencionadas. Embora existam várias modalidades de aprendizado federado, como a horizontal, vertical e a por transferência, este projeto propõe a construção de um modelo de aprendizado federado baseado em técnicas de clonagem de conhecimento de redes neurais com dados não-rotulados [8]. A hipótese é que combinando entradas aleatórias e as saídas obtidas pelos diferentes modelos espalhados nos dispositivos do aprendizado federado, é possível treinar um único modelo que consiga agregar todo o conhecimento extraído pelos demais sem necessidade de comunicações frequentes entre os nós e o modelo central e agregando de maneira mais lógica o conhecimento desenvolvido em cada nó.

Referências Bibliográficas

1. NGUYEN, G. T. et al. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review, v. 52, p. 77–124, 2019.
2. TULADHAR, A. et al. Building machine learning models without sharing patient data: A simulation-based analysis of distributed learning by ensembling. Journal of biomedical informatics, p. 103424, 2020.
3. ZHANG, T.; HE, Z.; LEE, R. B. Privacy-preserving machine learning through data obfuscation. ArXiv, abs/1807.01860, 2018.
4. MCMAHAN, H. B. et al. Communication-efficient learning of deep networks from decentralized data. In: International Conference on Artificial Intelligence and Statistics. [S.l.: s.n.], 2016.
5. ALEDHARI, M. et al. Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access, v. 8, p. 140699–140725, 2020.
6. SHAHID, O. et al. Communication efficiency in federated learning: Achievements and challenges. ArXiv, abs/2107.10996, 2021.
7. CUI, S. et al. Collaboration equilibrium in federated learning. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2021.
8. MOSAFI, I.; DAVID, E.; NETANYAHU, N. S. Stealing knowledge from protected deep neural networks using composite unlabeled data. 2019 International Joint Conference on Neural Networks (IJCNN), p. 1–8, 2019.