

Disciplina: INF16179 - Sistemas Distribuídos

Alunos: Thiago da Silva Meireles de Souza

Trabalho II

Metodologia

Este trabalho implementa um sistema distribuído descentralizado para a realização de um aprendizado federado hierárquico. É implementado um algoritmo de eleição de líder/coordenador a cada iteração/round até que se encerre o aprendizado. São usados um servidor agregador central, e dois grupos, podendo ter n clientes cada um, esses clientes por sua vez a cada round, fazem uma eleição para um agregador local e executam o aprendizado. Um round se dá pelos seguintes o fluxo:

1. Eleição do agregador local de cada grupo
2. *Treinamento local*
3. Envio dos pesos para o agregador local
4. *Agregação local*
5. Envio dos pesos para o agregador final
6. *Agregação final*
7. *Devolução dos pesos para os clientes*
8. Avaliação local
9. Envio da avaliação para o agregador local
10. *Agregação da avaliação local*
11. Envio dos pesos para avaliação final
12. *Agregação da avaliação final*
13. *Devolução do resultado da avaliação para os clientes*

Como forma de comunicação, é utilizado a troca de mensagens via publish/subscribe usando do broker EMQX, o qual foi executado localmente em container docker e configurado para permitir a troca de mensagens com até 10MB. Na implementação um único programa foi criado, contudo em sua inicialização são enviados diferentes parâmetros, que farão a distinção do modo de operação.

Os testes foram feitos com a realização do aprendizado em dois grupos, contendo cada um 4, 6, 8, e 10 clientes e um servidor agregador central, com um máximo de 25 rounds e meta de accuracy de 99,6%. A execução foi realizada por meio de um script responsável por realizar a chamada do programa na correta ordem e passando os devidos parâmetros.

Resultados Obtidos

A figura 1 e 2 apresenta o resultado obtido dos quatro aprendizados realizados, exibindo a accuracy e o número de rounds. Ambas as figuras apresentam os mesmos dados, contudo a figura 2 faz um recorde do round [2-25], permitindo uma melhor visualização para comparação dos resultados, o

aprendizado obteve accuracy acima de 98%. É possível notar que a velocidade de convergência demora um pouco mais quando o número de clientes aumenta, o que pode ser devido ao tamanho do dataset. Devido a necessidade de eleições constantes e sincronizações, o aprendizado hierárquico apresentou um tempo de execução maior, o que já era esperado devido à maior complexidade na comunicação e sincronização.

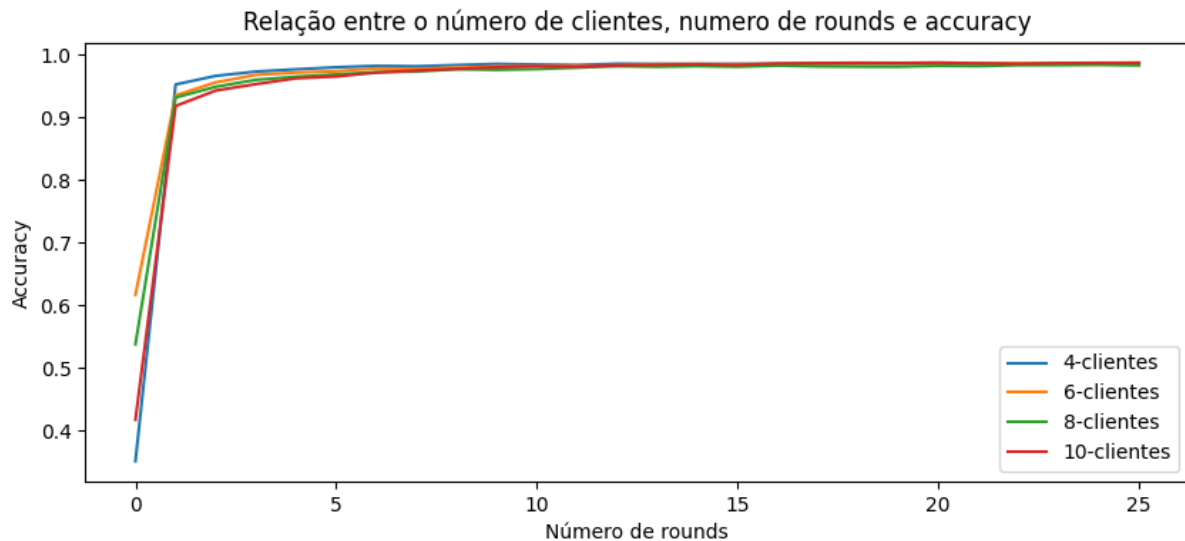


Figura 1

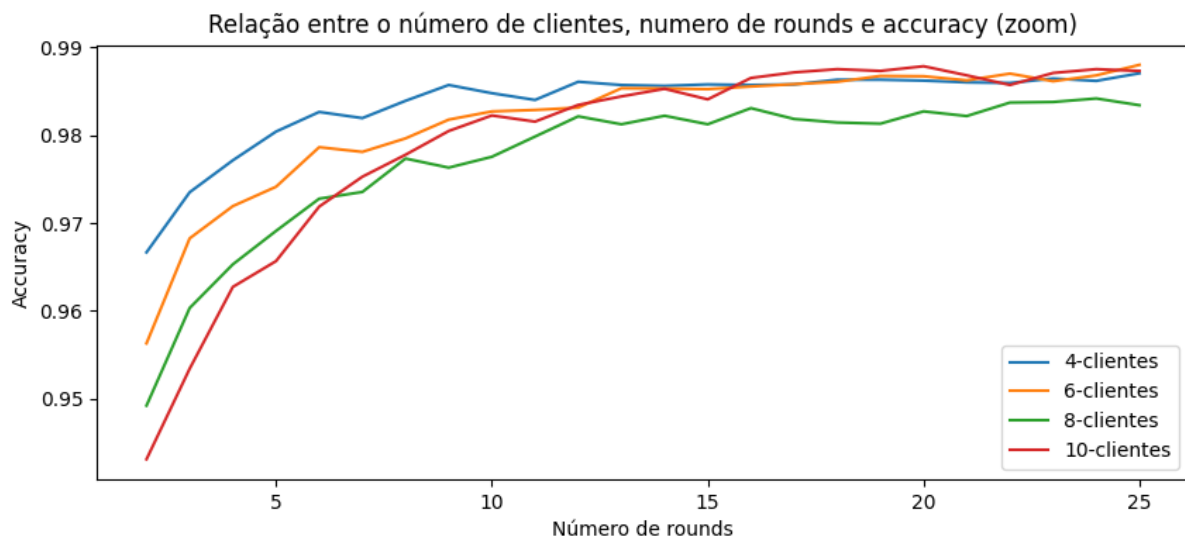


Figura 2

Discussões

A implementação do aprendizado federado hierárquico, embora tenha um leve aumento no tempo de execução, permite a flexibilização do aprendizado, contudo gera uma maior complexidade no processo, as eleições constantes e principalmente as sincronizações causam um aumento no tempo de execução, contudo esse modelo possui certa tolerância a falhas, tendo em vista que a

ausência, desde que não do agregador central, no sistema será transparente e não causará uma interrupção do aprendizado.