

A Survey on Distributed Machine Learning

Rapport

AKAMBA MANI Crescence Catherine

Departement Informatique
Faculté des sciences
Université de Yaoundé I

November 29, 2023



Table des matières

- 1 Introduction
- 2 Techniques d'augmentation du stockage
- 3 Architecture pour l'apprentissage automatique distribué
- 4 Quelques solutions d'apprentissage automatique distribué



Introduction

Au cours de ces dernières décennies, nous pouvons observer le développement remarquable de l'apprentissage automatique. Il a évolué des problèmes simples aux problèmes complexes. Les problèmes complexes nécessitent un grand nombre de données durant la phase d'entraînement. Ce qui peut causer une augmentation de la charge de travail, ainsi qu'un coût élevé du temps ainsi que de l'espace. Il devient donc indispensable de mettre en place des solutions pouvant résoudre ces problèmes.

L'apprentissage automatique distribué se présente comme une solution déterminante grâce à sa faculté à partitionner les problèmes en des sous-problèmes qui peuvent être résolus de façon parallèle sur plusieurs nœuds.



Techniques d'augmentation du stockage

Il existe deux façons différentes et complémentaires d'augmenter du stockage afin d'accélérer la charge du travail :

- Scaling-up

Ajout de plusieurs ressources dans une seule machine (noeud)

- GPUs Programmables
- GPGPU (General-purpose GPU)
- ASICs (Application Specific Integrated Circuits)

- Scaling-out

Ajout de plusieurs noeuds au système



Architecture pour l'apprentissage automatique distribué

L'architecture automatique distribuée est composée de trois couches dépendantes: apprentissage automatique , parallélisme et la topologie,

- ① Apprentissage automatique
- ② Parallélisme
 - Parallélisation des données
 - Parallélisation du modèle
- ③ Topologie
 - Système centralisée (ensembliste)
 - Système decentralisee (arbre)
 - Système decentralisee (parameter server)
 - Système distribué (Peer-to-peer)

Communication



Architecture pour l'apprentissage automatique distribué

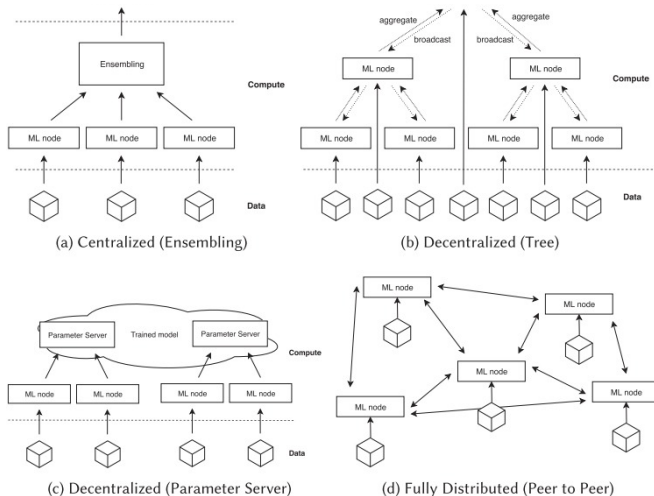


Fig. 3. Distributed machine learning topologies based on the degree of distribution.



Quelques solutions

L'architecture automatique distribuée est composée de trois couches dépendantes: apprentissage automatique , parallélisme et la topologie,

- MapReduce
- Apache spark
- Caffé 2
- ect...





Merci pour votre attention

