

计算通信理论论文阅读笔记01

517021911042 吴宁远

研究专题: Communication Cost

研究原因: 联合学习中的一项挑战是减少客户端与服务器之间的通信, 因为终端设备通常具有非常有限的通信带宽。

Standardized Federated Learning Process

Communication-Efficient Learning of Deep Networks from Decentralized Data

核心思想: The Federated-Averaging Algorithm

FederatedAveraging算法总共有三个基本的参数, C (0到1) 控制有多少比例的客户端参与优化, E 控制每一轮多少轮SGD需要在客户端运行, B 是每一轮的Mini-Batch的数目大小。另外, 假设总共有 K 个客户端。

- 每一轮都**随机选择**出 $\max(CK, 1)$ 个的客户端
- 对于**每个客户端**进行**Mini-Batch**的大小为 B , **轮数**为 E 的SGD更新
- 对于参数直接进行**加权平均** (这里的**权重**是每个**客户端的数据相对大小**)

文章对这里的最后一步进行了说明。之前有其他研究表明, 如何直接对参数空间进行加权平均, 特别是Non-Convex的问题, 会得到任意坏的结果。这篇文章里, 作者们对于这样的问题的处理是, 让**每一轮的各个客户端的起始参数值相同** (也就是**前一轮的全局参数值**)。这一步使得算法效果大幅度提高。

Communication-Efficient Federated Deep Learning with Asynchronous Model Update and Temporally Weighted Aggregation

具有异步模型更新和临时加权聚合功能的高效通信的联合深度学习

核心思想: 客户端异步学习; 服务器对本地模型进行时间加权聚合

Abstract: 本文通过在客户端上提出一种同步学习策略以及在服务器上对本地模型进行时间加权聚合, 提出了一种增强的联合学习技术。在异步学习策略中, 将深度神经网络的不同层分为浅层和深层, 并且深层的参数更新的频率比浅层的更新频率低。此外, 在服务器上引入了时间加权聚合策略, 以利用先前训练的局部模型, 从而提高中央模型的准确性和收敛性。

Introduction on Federated Learning

Four Step:

1. Parameters Download(Client)
2. Local Training(Client)
3. Trained Parameter upload(Client)
4. Model Aggregation(Server)

Challenges:

1. Unbalanced Data: Data Amount
2. Non-IID Data: Local datasets are not able to represent the overall data distribution
3. Massively Distributed Data
4. Unreliable Participating Clients

Performance Indicator:

Total Communication Cost

Works:

Reduce the communication cost and improve the accuracy of the central model

Main Contributions:

1. Asynchronous Strategy that aggregates and updates the parameters in the shallow and deep layers of DNNs at different frequencies.
2. Temporally weighted aggregation strategy

Related Work

Konecny:

Two ways to reduce communication costs:

1. Structured Updates
2. Sketched updates

FedAVG:

Same with 5.3

Asynchronous Model Update and Temporally Weighted Aggregation

Asynchronous Model Update

目的：降低联合学习的上传/下载的数据量（即降低通信成本）

核心思想：仅更新部分模型参数，浅层神经网络的参数更新频率远大于深层神经网络的参数更新频率。

Shallow layers: General Features(applicable to different datasets)

Deep layers: Specific Features

Shallow layers are useful to the general model.

Temporally Weighted Aggregation

目的：解决t-p轮和t轮参数更新重要性不同的问题

核心思想：最近被更新的模型参数应该拥有更大的模型权重

Framework

fig.5

Temporally Weighted Asynchronous Federated Learning

Algorithm 2: Server

Algorithm 3: Client

Client Component:

Step 1: 进行数据分批次

Step 2: 选择是对所有模型浅层/所有层进行训练

Step 3: 进行本地模型训练

Step 4: 返回模型参数

Server Component:

Step 1: 模型参数初始化，时间戳初始化

Step 2: 选择哪几个时间上传所有层/浅层参数

Step 3: 选择哪些Client上传参数，上传后更新时间戳

Step 4: 根据时间的距离进行参数的更新

Experimental Results and Analyses

Comparison

FedAVG: 基准算法

TWFL: 仅有异步模型更新

AFL: 仅有时间加权聚合

TWAF: 两者兼备

Important parameters

freq: 有几轮需要更新全部层信息

a: 用于标注时间距离

K: local clients的个数

m: 每轮选择的上传参数的client的个数

Settings on Datasets

目的: 反映non-IID、数据量不平衡以及massive distributed data等等问题

算法设计:

1) Handwritten Digit Recognition Using CNN

Nc: 控制每个本地数据集中的类数

Smin&Smax: 指定本地数据的大小范围

Labels: 指示相应任务中涉及的类的名称

2) Human Activity Recognition Using LSTM

Results and Analysis

1. freq: 所有层参数更新频率越小, 传输成本越小, 但是相应的准确率会降低;
2. a: 当以往轮的数据对于当前轮数据的影响越小, 准确率越高 (仅限于e与e/2的比较)
3. 随着客户端数量的增加, 所提出的算法相对于传统联合学习的优势将变得更加明显。
4. 有效降低了通信成本: 性能提升接近40%