Federated Multi-Task Learning联邦多任务学习

原創 静静不是万能的 2019-05-13 20:35

本论文翻译自原文Federated Multi-Task Learning

Abstract

Federated learning poses new statistical and systems challenges in training machine learning models over distributed networks of devices. In this work, we show that multi-task learning is naturally suited to handle the statistical challenges of this setting, and propose a novel systems-aware optimization method, MOCHA, that is robust to practical systems issues. Our method and theory for the first time consider issues of high communication cost, stragglers, and fault tolerance for distributed multi-task learning. The resulting method achieves significant speedups compared to alternatives in the federated setting, as we demonstrate through simulations on real-world federated datasets.

抽象

联合学习在分布式设备网络上训练机器学习模型方面带来了新的统计和系统挑战。 在这项工作中，我们表明多任务学习自然适合处理这种设置的统计挑战，并提出一种新的系统感知优化方法，MOCHA，它对实际系统问题是健壮的。 我们的方法和理论首次考虑了分布式多任务学习的高通信成本，落后者和容错问题。 与联合设置中的替代方案相比，所得到的方法实现了显着的加速，正如我们通过对真实联合数据集的模拟所证明的那样。

1 Introduction

Mobile phones, wearable devices, and smart homes are just a few of the modern distributed networks generating massive amounts of data each day. Due to the growing storage and computational power of devices in these networks, it is increasingly attractive to store data locally and push more network computation to the edge. The nascent field of federated learning explores training statistical models directly on devices [37]. Examples of potential applications include: learning sentiment, semantic location, or activities of mobile phone users; predicting health events like low blood sugar or heart attack risk from wearable devices; or detecting burglaries within smart homes [3, 39, 42]. Following [25, 36, 26], we summarize the unique challenges of federated learning below.

1简介

移动电话，可穿戴设备和智能家居只是现代分布式网络中的一小部分，每天都会产生大量数据。 由于这些网络中设备的存储和计算能力不断增长，因此在本地存储数据并将更多网络计算推向边缘变得越来越有吸引力。 新兴的联邦学习领域直接在设备上探索训练统计模型[37]。 潜在应用的示例包括：学习情感，语义位置或移动电话用户的活动; 预测可穿戴设备的低血糖或心脏病发作等健康事件; 或检测智能家居中的盗窃案[3,39,42]。 在[25,36,26]之后，我们总结了下面联邦学习的独特挑战。

1. Statistical Challenges: The aim in federated learning is to fit a model to data, {X1, . . . , Xm}, generated by m distributed nodes. Each node, t ∈ [m], collects data in a non-IID manner across the network, with data on each node being generated by a distinct distribution Xt ∼ Pt. The number of data points on each node, nt, may also vary significantly, and there may be an underlying structure present that captures the relationship amongst nodes and their associated distributions.

2. Systems Challenges: There are typically a large number of nodes, m, in the network, and communication is often a significant bottleneck. Additionally, the storage, computational, and communication capacities of each node may differ due to variability in hardware (CPU, memory), network connection (3G, 4G, WiFi), and

静 静静不是万能的

24小時熱門文章

最新文章

Federated Multi-Task Learning 聯邦多任務學習
opencv 學習筆記1
Deeplab圖像語義分割網絡-學習筆記
DeepLab_V3 Image Semantic Segmentation Network
ArcGIS Python 操作地圖的基礎入門筆記

最新評論文章

https://yachay.unat.edu.pe/blog/index.php?comment_area=format_blog&comment_component=blog&comment_co
linux以太網驅動總結

power (battery level). These systems challenges, compounded with unbalanced data and statistical heterogeneity, make issues such as stragglers and fault tolerance significantly more prevalent than in typical data center environments

1.统计挑战：联邦学习的目标是使模型适合数据，{X1 ,. 。 。 ，Xm}，由m个分布式节点生成。每个节点t∈[m]在网络上以非IID方式收集数据，每个节点上的数据由不同的分布Xt_Pt生成。每个节点上的数据点的数量nt也可以显着变化，并且可能存在捕获节点及其相关分布之间的关系的底层结构。
2.系统挑战：网络中通常存在大量节点m，并且通信通常是一个重要的瓶颈。另外，由于硬件（CPU，存储器），网络连接（3G，4G，WiFi）和电源（电池电平）的可变性，每个节点的存储，计算和通信容量可能不同。这些系统的挑战与不平衡的数据和统计异构性相结合，使得落后者和容错等问题比典型的数据中心环境更为普遍

In this work, we propose a modeling approach that differs significantly from prior work on federated learning, where the aim thus far has been to train a single global model across the network [25, 36, 26]. Instead, we address statistical challenges in the federated setting by learning separate models for each node, {w1, . . . , wm}. This can be naturally captured through a multi-task learning (MTL) framework, where the goal is to consider fitting separate but related models simultaneously [14, 2, 57, 28]. Unfortunately, current multi-task learning methods are not suited to handle the systems challenges that arise in federated learning, including high communication cost, stragglers, and fault tolerance. Addressing these challenges is therefore a key component of our work.

在这项工作中，我们提出了一种建模方法，它与联邦学习的先前工作有很大不同，迄今为止的目标是在整个网络中训练单一的全球模型[25,36,26]。 相反，我们通过学习每个节点的单独模型来解决联邦环境中的统计挑战{w1 ,.。 。 ，wm}。 这可以通过多任务学习（MTL）框架自然地捕获，其目标是同时考虑拟合单独但相关的模型[14,2,57,28]。 遗憾的是，当前的多任务学习方法不适合处理系统挑战在联邦学习中出现，包括高通信成本，落后者和容错。 因此，应对这些挑战是我们工作的重要组成部分。

1.1 Contributions

We make the following contributions. First, we show that MTL is a natural choice to handle statistical challenges in the federated setting. Second, we develop a novel method, MOCHA, to solve a general MTL problem. Our method generalizes the distributed optimization method COCOA [22, 31] in order to address systems challenges associated with network size and node heterogeneity. Third, we provide convergence guarantees for MOCHA that carefully consider these unique systems challenges and provide insight into practical performance. Finally, we demonstrate the superior empirical performance of MOCHA with a new benchmarking suite of federated datasets.

1.1贡献

我们做出以下贡献。 首先，我们表明MTL是处理联邦环境中统计挑战的自然选择。 其次，我们开发了一种新方法MOCHA来解决一般的MTL问题。 我们的方法推广了分布式优化方法COCOA [22,31]，以解决与网络规模和节点异构性相关的系统挑战。 第三，我们为MOCHA提供融合保证，认真考虑这些独特的系统挑战并提供洞察力实际表现。 最后，我们通过一个新的联邦数据集基准测试套件展示了MOCHA的卓越经验性能。

2 Related Work

Learning Beyond the Data Center. Computing SQL-like queries across distributed, low-powered nodes is a decades-long area of research that has been explored under the purview of query processing in sensor networks, computing at the edge, and fog computing [32, 12, 33, 8, 18, 15]. Recent works have also considered training machine learning models centrally but serving and storing them locally, e.g., this is a common approach in mobile user modeling and personalization [27, 43, 44]. However, as the computational power of the nodes within distributed networks grows, it is possible to do even more work locally, which has led to recent interest in federated learning.2 In contrast to our proposed approach, existing federated learning approaches [25, 36, 26, 37] aim to learn a single global model across the data.3 This limits their ability to deal with non-IID data and structure amongst the nodes. These works also come without convergence guarantees, and have not addressed practical issues of stragglers or fault tolerance, which are important characteristics of the federated setting. The work proposed here is, to the best of our knowledge, the first federated learning framework to consider these challenges, theoretically and in practice.

2相关工作

学习超越数据中心。跨分布式，低功耗节点计算类似SQL的查询是一个长达数十年的研究领域，已经在传感器网络中的查询处理，边缘计算和雾计算的范围内进行了探索[32,12,33,8 ，18,15]。最近的工作还集中考虑了训练机器学习模型，但是在本地服务和存储它们，例如，这是移动用户建模和个性化的常用方法[27,43,44]。然而，随着分布式网络内节点的计算能力的增长，可以做到均匀

更多的本地工作，这导致了最近对联邦学习的兴趣.2与我们提出的方法相比，现有的联邦学习方法[25,36,26,37]旨在学习单一的全球模型跨越数据.3这限制了它们在节点之间处理非IID数据和结构的能力。这些作品也没有收敛保证，也没有解决实际问题

落后者或容错，这是联邦设置的重要特征。据我们所知，这里提出的工作是第一个在理论上和实践中考虑这些挑战的联邦学习框架

近期，抖音上一款 “變身漫畫” 的特效刷爆“我的關注”，二次元漫畫樣式的畫風更是讓大家欲罷不能。從明星到路人，從大朋友到小朋友紛紛參與其中，抖音 App 中“變身漫畫”話題頁顯示約有 1851 萬餘人使用該道具。 如今，此類視頻 / 圖片

Yao 2020-07-14 12:03:51

深度学习_目标检测_FPN论文详解

FPN的創新點 多層特徵 特徵融合 解決了目標檢測中的多尺度問題，通過簡單的網絡連接改變，在基本不增加原有模型計算量的情況下，大幅度提升小物體（small object）的檢測性能。 在物體檢測裏面，有限計算量情況下，網絡的深

CV-GANRocky 2020-07-08 11:57:07

深度学习_目标检测_“YOLOv5”详解（持续更新）

YOLOv5可以方便的進行工程化部署： YOLOv5（PyTorch） −>ONNX−>CoreML−>iosYOLOv5（PyTorch） -> ONNX -> CoreML -> iosYOLOv5（PyTorch） −>ONNX

CV-GANRocky 2020-07-08 11:57:07

分类问题中的决策面画法 (直观理解plt.contour的用法)

摘要 通過分類問題中決策面的繪製過程直觀理解matplotlib中contour的用法，主要包括對 np.meshgrid 和plt.contour的直觀理解。 前言 分類問題中，我們習慣用2維的dmeo做例子，驗證算法的有效性。

张王李刘赵孙杨 2020-07-08 11:10:55

KNN算法 第二章 Pandas & sklearn 机器学习实战 Machine Learning in action

本專欄計劃藉助Pandas與sklearn重新實現書中的實戰案例。 k-近鄰算法1. KNN算法流程2. KNN改進約會網站的配對效果2.1 數據準備：從文本中解析數據2.2 數據可視化：散點圖2.3 數據處理：歸一化數值2.4

RPG_Zero 2020-07-08 11:09:44

[R]聚类算法:k-means模组

延伸<[Excel]k-means聚類算法的應用，以評價現有供應商的水平為例。>文章，同時恰巧在圖書館看到一本R語言機器學習書籍，因此正好可進一步瞭解如何用R語言來實現k-means算法和應用，一併將k-means模組建立起來，做為未來參

Learn-Share_HY 2020-07-08 10:38:53

python--内置函数

1、python內置函數： 類型轉換 數學運算 常用 int() max() all() range() help() float() min() any() set() format() long() sum() type()

沸点数据 2020-07-08 10:38:40

吴恩达机器学习课程思维导图

Github上黃博整理的吳恩達機器學習課程的資料，用xmind轉化成思維導圖，方便查看和記憶。 參考自：https://github.com/fengdu78/Coursera-ML-AndrewNg-Notes

阔岩 2020-07-08 09:20:14

机器学习笔记(七)--理解batch_dot函数

在keras中有batch_dot函數，用於計算兩個多維矩陣，官方註釋如下： def batch_dot(x, y, axes=None): """Batchwise dot product. `batch_dot`

LawGeorge 2020-07-08 09:20:14

置信学习：让样本中的“脏数据“原形毕露

在實際工作中，你是否遇到過這樣一個問題或痛點：無論是通過哪種方式獲取的標註數據，數據標註質量可能不過關，存在一些錯誤？亦或者是數據標註的標準不統一、存在一些歧義？特別是badcase反饋回來，發現訓練集標註的居然和badcase一樣？如下

hellozhxy 2020-07-08 09:16:27

simple faster rcnn解读 一

一:代碼框架和跑通simple faster rcnn遇到的問題 代碼選擇：https://github.com/chenyuntc/simple-faster-rcnn-pytorch； 本文主要是自己將代碼跑通中遇到的問題以及代碼解讀

charleswangzi 2020-07-08 09:02:08

决策树python实现（ID3 和 C4.5）

最近在看機器學習實戰，記錄一些不寫代碼，真的很難發現的問題。 ID3代碼見github ID3的問題： 1、從信息增益的計算方法來看，信息增益無法直接處理連續取值的的屬性數據，只能處理離散型的數據。 2、信息增益的計算方法需要對某

wf592523813 2020-07-08 08:32:14