Federated Learning & Distributional Shift in Healthcare

Ankit Pal (Monk)

Who Am I?

AI Research Engineer @ Saama AI Lab

- I work in Graph Learning and NLP field
- I am interested in
  - Representation Learning on Graphs and Manifolds
  - Privacy preserved Federated learning in Healthcare
  - Large Language Models(LLMs)
  - Biomedical signal processing
  - Spiking Neural networks

I also do :D

Cool Research | Trekking | Judo | Boxing | Skiing | Chess

Deep Learning

Goal of Deep Learning

Finding the function: model training

- Given one input sample pair (x_0, y_0), the goal of deep learning model training is to find a set of parameters w, to maximize the probability of outputting y_0 given x_0.

Given input: x_0

Maximize: $p(5|x_0, w)$

The importance of data for ML

> “The biggest obstacle to using advanced data analysis isn’t skill base or technology; it’s plain old access to the data.”

-Edd Wilder-James, Harvard Business Review

I am Data (In Thanos Voice)

"Data is the New Oil"

A SHIFT OF PARADIGM: FROM CENTRALIZED TO DECENTRALIZED DATA

The standard setting in Machine Learning (ML) considers a **centralized dataset processed in a tightly integrated system**

A SHIFT OF PARADIGM: FROM CENTRALIZED TO DECENTRALIZED DATA

But in the real world **data is often decentralized across many parties**

A SHIFT OF PARADIGM: FROM CENTRALIZED TO DECENTRALIZED DATA

But in the real world **data is often decentralized across many parties**

WHY CAN’T WE JUST CENTRALIZE THE DATA?

1. Sending the data may be too costly
   - Self-driving cars are expected to generate several TBs of data a day
   - Some wireless devices have limited bandwidth/power
2. Data may be considered too sensitive
   - We see a growing public awareness and regulations on data privacy
   - Keeping control of data can give a competitive advantage in business and research

HOW ABOUT EACH PARTY LEARNING ON ITS OWN?

1. The local dataset may be too small
   - Sub-par predictive performance (e.g., due to overfitting)
   - Non-statistically significant results (e.g., medical studies)

2. The local dataset may be biased
   - Not representative of the target distribution

So many problems with data

There must be a solution

A BROAD DEFINITION OF FEDERATED LEARNING

- **Federated Learning (FL)** aims to collaboratively train a ML model while keeping the data decentralized

initialize model

I am confused, Explain in simple terms

Let’s understand how Google uses Federated Learning

Let’s understand how Google uses Federated Learning

https://www.youtube.com/watch?v=gbRJPa9d-VU

Understood :)

Types of FL Learning

Horizontal Federated Learning

sex	age	spo2	comorbidities	symptoms	oxygen_req	icu_num_days
m	52	74	0	...	200	7
f	23	89	1	...	150	13
f	42	90	0	...	190	5

Hospital A

sex	age	spo2	comorbidities	symptoms	oxygen_req	icu_num_days
m	25	70	0	...	120	19
f	32	85	1	...	120	4
m	68	65	0	...	210	11

Hospital B

sex	age	spo2	comorbidities	symptoms	oxygen_req	icu_num_days
m	46	74	0	...	150	7
m	84	89	1	...	300	20
f	39	90	0	...	200	10

Hospital C

Types of FL Learning

Vertical Federated Learning

Hospital A

name	sex	age	spo2	comorbidities	symptoms	oxygen_req	icu_num_days
Person A	m	52	74	0	...	200	7
Person B	f	23	89	1	...	150	13
Person C	f	42	90	0	...	190	5

Fitness Tracking App

name	avg_active_mins	avg_rhr	avg_sleep	avg_emot
Person A	120	65	8h30m	happy
Person B	40	70	5h30m	uninterested
Person S	300	52	6h30m	uninterested

Types of FL Learning

Types of FL Learning

Types of FL Learning

Types of FL Learning

FL Learning Methods

Federated Aggregation

- Area of active research
- Simple approach: aggregate weights or gradients from local models.
- Secure aggregation: aggregate encrypted local updates and decrypt the result.
- Several caveats, discussed in the Challenges section.

FL Learning Methods

The Ingredients

- Centrally Coordinating Server
- A modelling and data processing utility
- A communication channel - we use websockets
- A medium to transfer local updates - we use Kafka
- Naive model averaging
- Tracking History

FL Learning Methods

Fig. 6. Machine Learning pipeline.

FL Learning Methods

The Recipe - Server

```
class Server:
    def __init__(self):
        pass

    async def connect(self, sid, environ):
        pass
```

FL Learning Methods

```
class Server:
    def __init__(self):
        pass

    async def connect(self, sid, environ):
        # connect with nodes, start training on min nodes

    async def start_round(self):
        # start a training round and send global model

    async def fl_update(self, sid, data):
        # receive ack for updates

    def consume_updates(self):
        # consume updates when all updates are received
```

FL Learning Methods

```
class Server:
....
   async def fl_update(self, sid, data):
       # receive ack for updates

   def consume_updates(self):
       # consume updates when all updates are received

   def aggregate(self, client_mapped_weights):
       # aggregate weights for layers with trainable weights

   def evaluate(self, aggregated_weights):
       # Evaluate on a holdout set

   def store_history(self):
       # Store federated losses across rounds
```

FL Learning Methods

The Recipe - Client

```
class Node:
    def __init__(self):
        pass

    async def connect(self, sid, environ):
        # Connect to the server
```

FL Learning Methods

```
class Node:
    def __init__(self, address, partition, client, epochs):
        pass

    def connect(self):
        # Connect to server

    def connection_received(self):
        # Get ack from server
```

FL Learning Methods

```
class Node:

    def start_training(self, _model):
        # get model from json
        # compile model
        # fit
        # evaluate
        # send updates
        pass

    def fit(self, model):
        pass

    def send_updates(self, loss):
        # encode individual layers as b64 strings
        # produce over the updates topic on a Kafka server
        pass
```

FL Learning Methods

```
class Node:


    def end_session(self, data):
        # get latest model weights
        # update local model
        # Clean up as necessary
        pass

    def disconnect(self):
        # Disconnect signal from server
```

FL Frameworks

Federated Learning Frameworks

Introducing TensorFlow Federated

TensorFlow Federated enables developers to express and simulate federated learning systems. Pictured here, each phone trains the model locally (A). Their updates are aggregated (B) to form an improved shared model (C).

```
# Load simulation data.
source, _ = tff.simulation.datasets.emnist.load_data()
def client_data(n):
  dataset = source.create_tf_dataset_for_client(source.client_ids[n])
  return mnist.keras_dataset_from_emnist(dataset).repeat(10).batch(20)

# Wrap a Keras model for use with TFF.
def model_fn():
  return tff.learning.from_compiled_keras_model(
      mnist.create_simple_keras_model(), sample_batch)

# Simulate a few rounds of training with the selected client devices.
trainer = tff.learning.build_federated_averaging_process(model_fn)
state = trainer.initialize()
for _ in range(5):
  state, metrics = trainer.next(state, train_data)
  print (metrics.loss)
```

Source: https://blog.tensorflow.org/2019/03/introducing-tensorflow-federated.html

FL Frameworks

Federated Learning Frameworks

README.md

TensorFlow Privacy

This repository contains the source code for TensorFlow Privacy, a Python library that includes implementations of TensorFlow optimizers for training machine learning models with differential privacy. The library comes with tutorials and analysis tools for computing the privacy guarantees provided.

The TensorFlow Privacy library is under continual development, always welcoming contributions. In particular, we always welcome help towards resolving the issues currently open.

Source:https://blog.tensorflow.org/2019/03/introducing-tensorflow-privacy-learning.html

FL Frameworks

Federated Learning Frameworks

A generic framework for privacy preserving deep learning

Théo Ryffel*
Imperial College London
tr17@ic.ac.uk

Andrew Trask*
DeepMind
University of Oxford
liamtrask@gmail.com

Morten Dahl*
mortendahlcs@gmail.com

Bobby Wagner*
Case Western Reserve University
bobbywagner@case.edu

Jason Mancuso*
jason@manc.us

Daniel Rueckert
Imperial College London
dr@ic.ac.uk

Jonathan Passerat-Palmbach
Imperial College London
jpassera@ic.ac.uk

Source:https://github.com/OpenMined/PySyft

Federated Learning in Healthcare

Federated Learning is the Future to share private data

FL in Healthcare

Federated Learning in Healthcare (Prediction)

Research setting [**Liu et al. 2018**]

ICU patients data from 208 critical care units b/w 2014 and 2015 (**eICU** by Philips)

58 hospitals with 1,264,89 ICU admissions

Input: medications taken first 24 hrs
Binary feature vector [1400x1]

Primary outcome: Patient mortality

Data description

Table 1: Study cohort (count by admission)

Information	Alive	Deceased
Age		
0-18	4269	403
18-60	48777	1945
>60	66867	4558
Gender		
Female	54286	3094
Male	65286	3789
Medical information		
Length of stay in hrs, mean(std)	65.5(92.6)	98.3(162.5)
Number of drug started in the first 24 hours	12.9(10.0)	13.6(9.2)

FL in Healthcare

Federated Learning in Healthcare (Prediction)

eICU Results [Liu et al. 2018]

Table 2: Performance of centralized, original federated and Federated-autonomous learning

Training method	AUCROC	AUCPR
Centralized learning	0.79	0.21
Original federated learning	0.75	0.16
Federated autonomous deep learning (FADL)	0.79	0.23

FL in Healthcare - Examples

JOURNAL ARTICLE

Federated learning improves site performance in multicenter deep learning without data sharing

Karthik V Sarma, Stephanie Harmon, Thomas Sanford, Holger R Roth, Ziyue Xu, Jesse Tetreault, Daguang Xu, Mona G Flores, Alex G Raman, Rushikesh Kulkarni ... Show more

Journal of the American Medical Informatics Association, Volume 28, Issue 6, June 2021, Pages 1259–1264, https://doi.org/10.1093/jamia/ocaa341

Published: 04 February 2021 **Article history**

Federated learning for computational pathology on gigapixel whole slide images

Ming Y. Lu [a, c, 1], Richard J. Chen [a, b, c, 1], Dehan Kong [a], Jana Lipkova [a, c], Rajendra Singh [e], Drew F.K. Williamson [a, c], Tiffany Y. Chen [a, c], Faisal Mahmood [a, c, d, f]

Show more

Federated Learning for Multi-Center Imaging Diagnostics: A Study in Cardiovascular Disease

Akis Linardos, Kaisar Kushibar, Sean Walsh, Polyxeni Gkontra, Karim Lekadir

Deep learning models can enable accurate and efficient disease diagnosis, but have thus far been hampered by the data scarcity present in the medical world. Automated diagnosis studies have been constrained by underpowered single-center datasets, and although some results have shown promise, their generalizability to other institutions remains questionable as the data heterogeneity between institutions is not taken into account. By allowing models to be trained in a distributed manner that preserves patients' privacy, federated learning promises to alleviate these issues, by enabling diligent multi-center studies. We present the first federated learning study on the modality of cardiovascular magnetic resonance (CMR) and use four centers derived from subsets of the M\&M and ACDC datasets, focusing on the diagnosis of hypertrophic cardiomyopathy (HCM). We adapt a 3D-CNN network pretrained on action recognition and explore two different ways of incorporating shape prior information to the model, and four different data augmentation set-ups, systematically analyzing their impact on the different collaborative learning choices. We show that despite the small size of data (180 subjects derived from four centers), the privacy preserving federated learning achieves promising results that are competitive with traditional centralized learning. We further find that federatively trained models exhibit increased robustness and are more sensitive to domain shift effects.

Saama Research in FL

Paper Title : **Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges**
Data Category : **Text data, Tabular data**
Accepted at : **Association for Computing Machinery (ACM) 2022**

Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges

MADHURA JOSHI*, ANKIT PAL*, and MALAIKANNAN SANKARASUBBU,
Saama AI Research, India

Federated learning is the process of developing machine learning models over datasets distributed across data centers such as hospitals, clinical research labs, and mobile devices while preventing data leakage. This survey examines previous research and studies on federated learning in the healthcare sector across a range of use cases and applications. Our survey shows what challenges, methods, and applications a practitioner should be aware of in the topic of federated learning. This paper aims to lay out existing research and list the possibilities of federated learning for healthcare industries.

CCS Concepts: • **Security and privacy**; • **Computing methodologies** → **Artificial intelligence**;;

Additional Key Words and Phrases: Federated learning, GDPR, transfer learning

ACM Reference format:
Madhura Joshi, Ankit Pal, and Malaikannan Sankarasubbu. 2022. Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges. *ACM Trans. Comput. Healthcare* 3, 4, Article 40 (October 2022), 36 pages.
https://doi.org/10.1145/3533708

40

1 INTRODUCTION
In the last few years, digital healthcare data has grown significantly. At the same time, recent breakthroughs in **deep learning (DL)** have been used in a variety of current medical data processes, including automatic disease diagnosis [72, 91], classification, biomedical data analysis, Question Answering in the medical domain [88], and segmentation [67, 76]. These methods hold immense promise and innovation in this field. In the coming future, the advancement of these methods will refine health care systems and improve medical practices worldwide.

Diagnostic tools, **machine learning (ML)** based healthcare solutions, and models must be exposed to a wide variety of cases and data that cover a full range of possible anatomies to capture more informative patterns in the medical data. It is well known that data from a single source can be significantly biased by the equipment, demographics, and acquisition protocol. Therefore, training a model on data from a single source would skew its prediction performance towards the population. Moreover, it is computationally expensive and time-consuming.

Training models in a parallelized manner [30] and within small batches can mitigate a few of these challenges. However, though this approach addresses computation challenges, it does not necessarily preserve the privacy of data. Clinical research often involves studies from a large amount of data collected from various sources. Health institutions, individuals, insurance companies, and the pharmaceutical industry all have access to medical data.

Key Contributions

- Demonstrate the components of the federated learning setup and discuss the communication architecture and building blocks of a federated learning system
- Examine the various challenges that a federated learning setup faces in terms of privacy, data, and communication in the healthcare system
- Survey existing works on federated learning in the health sector and propose a comprehensive list of applications classified into prognosis, diagnosis, and clinical workflow

https://dl.acm.org/doi/pdf/10.1145/3533708

Distribution shift Issues

Fig. 8. Applications of federated learning in healthcare.

Distribution shift Issues

Fig. 4.7.1 Training data for distinguishing cats and dogs.

At test time we are asked to classify the images in Fig. 4.7.2.

Fig. 4.7.2 Test data for distinguishing cats and dogs.

Distribution shift Issues

Paper Title : CLIFT : Analysing Natural Distribution Shift on Question Answering Models in Clinical Domain
Data Category : Language, Text data
Accepted at : Conference on Neural Information Processing Systems, Robustseq 2022

CLIFT : Analysing Natural Distribution Shift on Question Answering Models in Clinical Domain

Ankit Pal
Saama AI Research
Chennai, India
ankit.pal@saama.com

Abstract

This paper introduces a new testbed CLIFT (Clinical Shift) for the clinical domain Question Answering task. The testbed includes 7.3k high-quality question-answering samples to provide a diverse and reliable benchmark. We performed a comprehensive experimental study and evaluated several QA deep-learning models under the proposed testbed. Despite impressive results on the original test set, the performance degrades when applied to new test sets, which shows the distribution shift. Our findings emphasize the need for and the potential for increasing the robustness of clinical domain models under distributional shifts. The testbed offers one way to track progress in that direction. It also highlights the necessity of adopting evaluation metrics that consider robustness to natural distribution shifts. We plan to expand the corpus by adding more samples and model results. The full paper and the updated benchmark are available at openlifescience-ai.github.io/clift

1 Introduction

Digital healthcare data has significantly expanded during the last several years. Question Answering (QA) in the clinical domain is a particular task involved with clinical texts in Electronic Health Records (EHR). Clinical notes & documents in EHRs are frequently used by medical practitioners to find information that can aid in decision-making. [22] Recent studies using neural QA models [20, 3, 4] have shown promising outcomes. As a result, significant effort has been invested in developing question-and-answer algorithms & datasets in this field. The objective of developing clinical domain question-answering systems is not to achieve the highest possible accuracy. Instead, the goal is to generalize the model on new clinical data from various regions, hospitals, and organizations.

Key Contributions

- We are proposing five new test datasets. The testbed covers different clinical domain diseases and multiple sub-topics.
- Evaluate the diverse reasoning abilities of clinical models over a vast array of disease topics and subtopics.
- Detailed statistics and fine-grained evaluation of natural distribution shift are provided in this study

https://nips.cc/media/PosterPDFs/NeurIPS%202022/58229.png

Distribution shift Issues

Paper Title : **CLIFT : Analysing Natural Distribution Shift on Question Answering Models in Clinical Domain**
Data Category : **Language, Text data**
Accepted at : **Conference on Neural Information Processing Systems, Robustseq 2022**

Figure 1: Comparison of models' F1 scores on emrQA and our proposed test datasets. The result demonstrates that despite impressive results on the training dataset when applied to new test sets, the model's performance degrades, which shows the distribution shift

$$l_Q(M) - l_{Q'}(M) = \underbrace{(l_Q(M) - l_{\mathcal{P}}(M))}_{\text{Adaptivity gap}} + \underbrace{(l_{\mathcal{P}}(M) - l_{\mathcal{P}'}(M))}_{\text{Distribution gap}} + \underbrace{(l_{\mathcal{P}'}(M) - l_{Q'}(M))}_{\text{Generalization gap}}$$

Distribution shift Issues

Figure 4: Heat map of all layers' tokens & attributions for the start position prediction.

Distribution shift Issues

Paper Title : **CLIFT : Analysing Natural Distribution Shift on Question Answering Models in Clinical Domain**
Data Category : **Language, Text data**
Accepted at : **Conference on Neural Information Processing Systems, Robustseq 2022**

Model	emrQA F1	Clift F1	Gap (δ)
BERT-ClinicalQA	88.28	17.31	70.97
BioBert-B	87.55	17.38	70.17
BioClinicalBert	86.98	17.67	69.32
BlueBert-MIMIC	89.16	19.20	69.96
Electra-Squad-L	86.98	19.44	67.55
BlueBert-PubMed	86.98	19.81	67.18
DistilBert-squad-B	85.68	20.45	65.23
PubMedBERT	91.72	21.95	69.77
BioBert-squad-L	90.95	22.53	68.42
Bert-Large	88.81	23.25	65.56
Biomed-Roberta	89.76	23.59	66.18
Roberta-squad-B	89.84	24.99	64.85

Table 3: Performance of different models on the emrQA Testset and new Medication Testset

Model	emrQA F1	Clift F1	Gap (δ)
PubMedBERT	91.72	21.35	70.37
Biomed-Roberta	89.76	22.72	67.04
BioBert-B	87.55	23.37	64.18
Electra-Squad-L	86.98	24.09	62.89
BioClinicalBert	86.98	24.51	62.47
BlueBert-MIMIC	89.16	25.50	63.67
BioBert-squad-L	90.95	25.87	65.08
Bert-Large	88.81	26.17	62.63
BERT-ClinicalQA	88.28	27.04	61.24
DistilBert-squad-B	85.68	27.10	58.58
BlueBert-PubMed	86.98	27.54	59.44
Roberta-squad-B	89.84	29.90	59.94

Table 4: Performance of different models on the emrQA Testset and new Heart Testset

Model	emrQA F1	Clift F1	Gap (δ)
Biomed-Roberta	89.76	25.54	64.22
PubMedBERT	91.72	26.86	64.86
DistilBert-squad-B	85.68	29.19	56.49
BioBert-B	87.55	30.31	57.24
BERT-ClinicalQA	88.28	32.45	55.83
BioClinicalBert	86.98	33.46	53.52
Roberta-squad-B	89.84	33.73	56.11
BlueBert-MIMIC	89.16	36.49	52.67
BlueBert-PubMed	86.98	36.51	50.47
BioBert-squad-L	90.95	36.56	54.40
Electra-Squad-L	86.98	36.90	50.08
Bert-Large	88.81	37.31	51.49

Table 5: Performance of different models on the emrQA Testset and new Cancer Testset

Figure 4: Heat map of all layers' tokens & attributions for the start position prediction.

Model	emrQA F1	Clift F1	Gap (δ)
DistilBert-squad-B	85.68	12.62	73.07
BioBert-B	87.55	12.89	74.66
BERT-ClinicalQA	88.28	13.00	75.28
BlueBert-MIMIC	89.16	13.60	75.56
BlueBert-PubMed	86.98	14.07	72.91
BioClinicalBert	86.98	14.20	72.78
BioBert-squad-L	90.95	14.49	76.46
PubMedBERT	91.72	14.65	77.08
Electra-Squad-L	86.98	14.81	72.17
Bert-Large	88.81	16.92	71.88
Biomed-Roberta	89.76	17.31	72.46
Roberta-squad-B	89.84	19.88	69.96

Table 6: Performance of different models on the emrQA Testset and new Obesity Test

Model	emrQA F1	Clift F1	Gap (δ)
BioBert-B	87.55	32.57	54.98
BERT-ClinicalQA	88.28	34.31	53.97
Electra-Squad-L	86.98	35.13	51.85
DistilBert-squad-B	85.68	36.18	49.51
BioClinicalBert	86.98	36.90	50.08
Biomed-Roberta	89.76	38.38	51.39
BlueBert-PubMed	86.98	40.11	46.87
Bert-Large	88.81	42.84	45.97
BioBert-squad-L	90.95	43.36	47.60
PubMedBERT	91.72	44.75	46.97
Roberta-squad-B	89.84	45.16	44.68
BlueBert-MIMIC	89.16	45.34	43.82

Table 7: Performance of different models on the emrQA Testset and new Smoke Testset

```
{
    "data": {
        "Clift_data": [
            {
                "Answer": "The patient is 49 years old.",
                "Question": "What is the patient's age?",
            },
            {
                "Answer": "The patient has a history of "
                "hypertension, hypercholesterolemia, and "
                "eosinophilia.",
                "Question": "What is the patient's medical history?",
            },
            {
                "Answer": "The two-view chest x-ray showed that the "
                "right pneumonectomy space is mostly "
                "fluid-filled, with a prominent air-fluid "
                "level at the level of the aortic arch.",
                "Question": "What was the result of the patient's " "chest x-
ray?",
            },
        ],
        "context_id": "29a63316-4865-4454-970e-47ae5b273d4a",

        "paragraph": "xxln is a 49 yo gentleman with a biopsy-proven right "
        "lower lobe squamous cell carcinoma, metastatic to the "
        "n1 modes but with negative mediastinoscopy. past "
        "medical history: psychiatric history: -patient has "
        "never been in psychiatric tx. he does not have a "
        "psychiatrist or therapist, has never been "
        "hospitalized, denies suicidal attempt, and "
        "homicidal\\\\/assaultive behavior. . past medical "
        "history -htn -hypercholesteremia -eosinophilia - per "
        "pcp, w\\\\/up for strongyloides - prescribed tx, "
        "unsure of compliance . substance abuse history "
        "-smoking cigarrettes - quit smoking 6 years ago -etoh "
        "- some weekends; 4-5 beers; denies blackouts, w\\\\/d "
        "sx, and detox -denies iv and illicit drug use family "
        "history: non- contributory physical exam: general: "
        "well appearing male in nad. heent: unremarkable chest: "
        "right thoracotomy incision well approx. mild erythema, "
        "no drainage. cor: rrr s1, s2 abd: soft, round, nt, nd, "
        "+bs extrem: no c\\\\/c\\\\/e neuro: intact pertinent "
        "results: two-view chest xxdate comparison: xxdate. "
        "indication: status post pneumonectomy. patient is "
        "fully upright on current study and was likely "
        "semi-upright on the most recent study, limiting "
        "comparison. the right pneumonectomy space is mostly "
        "fluid filled, with a prominent air-fluid level "
        "demonstrated at the level of the aortic arch, "
        "corresponding to the sixth posterior right rib level.",
    }
}
```

Thank you for your attention!