A Performance Evaluation of Federated Learning algorithms

Adrian Nilsson, Simon Smith, Gregor Ulm, Emil Gustavsson, Mats Jirstrand

Presented by Simon Smith at DIDL’18

2018-12-10
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.

- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020\(^1\)

\(^1\) Hitachi, “The internet on wheels and Hitachi, Ltd.”, 2015
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.

- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020\(^1\)

\(^1\)Hitachi, “The internet on wheels and Hitachi, Ltd.”, 2015
What is Federated Learning (FL)?

- Distributed machine learning
  - Communicate a model, not data.

- Massive number of clients
  - Slow, unreliable network
  - 250M connected vehicles by 2020\(^1\)

- GDPR

\(^1\) Hitachi, “The internet on wheels and Hitachi, Ltd.”, 2015
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach
What have we done!?

- Implemented and compared three FL algorithms
- Compared with fully centralized approach
FedAvg

$w_{t+1} = \frac{1}{\mu} \sum_{k \in S_t} \frac{n_k}{\mu} w_{t+1}^k$

$\mu = \sum_{k \in S_t} n_k$

Server

C=0.4

Adrian Nilsson & Simon Smith
FedAvg

Server

$C = 0.4$

\[
\begin{align*}
  w_{t+1} &= \sum_{k \in S_t} \frac{n_k}{\mu} w^k_{t+1} \\
  \mu &= \sum_{k \in S_t} n_k
\end{align*}
\]
FSVRG

Server

\[ w_0 \]

\[ w_0 \]

\[ w_0 \]

\[ w_0 \]
FSVRG

Server

$\nabla w_0^1$

$\nabla w_0^2$

$\nabla w_0^3$

$\nabla w_0^4$

$\nabla w_0^5$
FSVRG

Server

$\nabla w_0 \quad \nabla w_0 \quad \nabla w_0 \quad \nabla w_0 \quad \nabla w_0$
FSVRG

Server

$w_1^1$  $w_1^2$  $w_1^3$  $w_1^4$  $w_1^5$
CO-OP

Age filters: $b_l, b_u$

Server age: $a$

$\iff b_l \leq a - a_k \leq b_u$

Client age: $a_k$
### Algorithms

<table>
<thead>
<tr>
<th></th>
<th>Synchronous</th>
<th>Opt. Algorithm</th>
<th>New hyperparam.</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>FedAvg</td>
<td>✔️</td>
<td>SGD</td>
<td>$C, E$</td>
<td>$C$ – fraction of clients</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$E$ – epoch before upload</td>
</tr>
<tr>
<td></td>
<td>✔️</td>
<td>SVRG</td>
<td>$h$</td>
<td></td>
</tr>
<tr>
<td>FSVRG</td>
<td></td>
<td></td>
<td>$b_l, b_u$</td>
<td>“Age filters” - Mitigate staleness</td>
</tr>
<tr>
<td>CO-OP</td>
<td></td>
<td>SGD</td>
<td>$b_l, b_u$</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation approach

MNIST digit recognition

Feed-forward ANN with 2 hidden layers

100 clients

Image by Josef Steppan [CC BY-SA 4.0], from Wikimedia Commons
Evaluation approach

IID & non-IID partitionings

Client 1, IID

Client 1, non-IID
Evaluation approach

▪ Hyperparameter search
  – Learning rate, decay, epochs, batch size, global step size, age filter

▪ Cross-validation
  – We allow 10,000 uploads from 100 simulated clients
Bayesian comparisons

- x-axis shows mean difference in accuracy between A and B

- Region of practical equivalence (rope)

- Area is interpreted as a probability.
Benchmarking on IID data
Results – Federated Learning IID

FedAvg vs CO-OP

FedAvg vs FSVRG

CO-OP vs FSVRG
Results – Federated Learning IID vs Centralized Learning

FedAvg

FSVRG

CO-OP
Benchmarking on non-IID data
Results – FL non-IID

FedAvg vs CO-OP

FedAvg vs FSVRG

CO-OP vs FSVRG
Results – FL non-IID

CO-OP vs FSVRG

FedAvg vs CO-OP
FedAvg vs FSVRG

Adrian Nilsson & Simon Smith
Results – FL non-IID vs Centralized Learning

FedAvg
FSVRG
CO-OP
Allowing more communication for FSVRG and CO-OP
Results – FL non-IID; more uploads

FedAvg vs CO-OP (×5)

FedAvg vs FSVRG (×10)

CO-OP (×5) vs FSVRG (×10)
Results – FL non-IID; more uploads vs Centralized Learning

FSVRG ($\times 10$)

CO-OP ($\times 5$)
Practical considerations

- **FSVRG**
  - Requires more communication per global update

- **CO-OP**
  - Age filters are difficult to tune
Future work

- Evaluate on multiple datasets
- Examine unevenly distributed data
  - i.e. a few cars hold most of the data
- New algorithms
Algorithm + "Federated Learning"

Hits on Google Scholar (from 2016)

- 2016: 13
- 2017: 161
- 2018: 358
What has happened since?

- Non-IID is still an issue [1,2]
- More privacy [3,4]
- New Algorithms
  - Asynchronous FL [5]
  - Federated Kernelized Multi-task Learning [7]
Thank you
References

(Non-IID)

(Policy)

(Algorithms)

Adrian Nilsson & Simon Smith