HKUST COMP 6211G:
Lecture 1: An Introduction to Federated Learning

Instructor: Qiang Yang

https://www.fedai.org/
Course Descriptions

• The world’s first course on Federated Learning.

• Concepts, techniques and systems about Federated Learning (FL), an emerging, privacy-preserving machine learning paradigm will be introduced.

• Students are also expected to:
  • Read and present research papers on federated learning, and
  • Carry out a research project on federated learning

• Course website:
  • Canvas, and also, [https://ising.cse.ust.hk/fl/](https://ising.cse.ust.hk/fl/)
  • Attend via Zoom on 10:30 a.m, TuTh: [https://hkust.zoom.us/j/92192958807](https://hkust.zoom.us/j/92192958807)
Prerequisites

• Students should have basic knowledge on:
  • Machine learning (e.g. optimization, decision trees, neural networks...)
  • Probability and linear algebra
  • Programming, e.g. Python
Topics Covered

• Backgrounds: Privacy-preserving machine learning, distributed machine learning...
• Federated Learning Concepts: Horizontal federated learning, vertical federated learning, federated transfer learning...
• Applications of FL: CV, Recommendation, Healthcare, ...
• Advances and Open Problems

• Reference textbook:
  • ISBN: 978-1681736983
Instructors

- Professor Qiang Yang, qyang@cse.ust.hk
- Professor Kai Chen, kaichen@cse.ust.hk
- Professor Yangqiu Song, yqsong@cse.ust.hk
- Dr. Lixin Fan, Dr. Yang Liu from WeBank

- TAs:
  - Zhenghang Ren, zrenak@connect.ust.hk
  - Yilun Jin, yilun.jin@connect.ust.hk
  - Tianjian Chen, tchenay@connect.ust.hk
- For questions regarding course logistics, please direct to Zhenghang and Yilun.
The grades will consist of four parts.

<table>
<thead>
<tr>
<th>Content</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation</td>
<td>5%</td>
</tr>
<tr>
<td>Mid-term exam</td>
<td>30%</td>
</tr>
<tr>
<td>Paper presentation</td>
<td>15%</td>
</tr>
<tr>
<td>Project (proposal &amp; final presentation)</td>
<td>50%</td>
</tr>
</tbody>
</table>
Timetable (tentative)

- The tentative timetable of this course is

<table>
<thead>
<tr>
<th>Week No.</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
<td>Lectures</td>
</tr>
<tr>
<td>4</td>
<td>Lab tutorial and Mid-term</td>
</tr>
<tr>
<td>5-7, 9-11</td>
<td>Paper Presentation</td>
</tr>
<tr>
<td>8</td>
<td>Project Proposal Presentation</td>
</tr>
<tr>
<td>12, 13</td>
<td>Final Project Presentation</td>
</tr>
</tbody>
</table>
Grading Details

• Mid-term:
  • Scheduled in week 4, covering lectures in week 1-3.
• Paper Presentation (From week 5):
  • A defense-offense style of presentation is adopted.
  • (Tentative) Students are expected to form groups of 2 for presentation. For each paper, one group will ‘defend’ it and another will ‘offend’ it.
  • We expect group formulations to be submitted on Feb. 16.
  • Paper assignments and presentation guidelines will be finalized before Feb. 18.
Final Project

• The final project aims to let students explore open research problems on federated learning.
  • Some topics are available on Canvas and https://ising.cse.ust.hk/fl/index.html?project_assignments, but you are also free to choose your own.
• Students are expected to form groups of 3-5.
• An oral proposal presentation (week 8), a final presentation (week 12, 13) and a conference-style written paper are expected.
Next Challenge: Data Are Fragmented, Data Silos
Background

- Increasingly strict laws on data protection:
  - GDPR of EU, 2018
  - CCPA of USA, 2018
  - Cyber Security Law of China, 2017

- Growing concern on user privacy and data security

- Data exist in the form of isolated silos.

- Federated learning can be a solution! [McMahan’16, Yang’19]

Reference:
[1] DLA Piper, Data Protection Laws of the World
   https://www.dlapiperdataprotection.com/
2. Data Sharing Among Parties: Difficult, Impossible, Illegal or Immoral

- Medical clinical trial data cannot be shared (by R. Stegeman 2018 on Genemetics)
- Our society demands more control on data privacy and security
  - GDPR, Government Regulations
  - Corporate Security and Confidentiality Concerns
  - Data privacy concerns
China’s Data Cyber Security Law

- Enacted in 2017
- Requires that Internet businesses must not leak or tamper with the personal information
- When conducting data transactions with third parties, they need to ensure that the proposed contract follow legal data protection obligations.
- More to come...

From Report by KPMG 2017
Facebook finally rolls out privacy tool for your browsing history

Google strengthens Chrome’s privacy controls

Top Microsoft exec says online privacy has reached 'a crisis point'
Challenges to AI: small data and fragmented data, Non-iid, Non-balanced, non-cooperative or malicious, dirty data, incomplete data, outdated data...

Low Security in Data Sharing
Lack of Labeled Data
Segregated Datasets

Over 80% of the world's enterprise information are in data silos!
Federated Learning

- Move models instead of data; Data usable but not visible.
- Multi-party model learning without exchanging data.

Scenarios:
- Collaboration among edge devices,
- Collaboration among organizations,
- Collaboration among departments within one organization.

Also known as:
- Federated Machine Learning
- Collaborative Machine Learning
- Federated Deep Learning
- Federated Optimization
- Privacy-preserving Machine Learning
- Geo-distributed Machine Learning
- Geo-distributed Deep Learning
- Multi-party Learning

References:
Definition of Federated Learning

- **Definition**
  - Multiple parties, each of which owns some data, collaborate to jointly train a machine learning model.
  - During training, no data held by each party will leave that party.
  - Information necessary to model training will be transferred under protection (e.g. encryption), to ensure that no party can re-engineer the data owned by others.
  - The performance of the resulting model should be a good approximation of the ideal model, built with all data transferred to a single party.

- **Categorization:**
  - Horizontal Federated Learning (Chapter 4)
  - Vertical Federated Learning (Chapter 5)
  - Federated Transfer Learning (Chapter 6)
  - Federated Reinforcement Learning (Chapter 9)

---

**Data Isolation:**
Raw data are not leaked or transferred.

**Lossless:**
Same accuracy as all data

**Equality:**
Participants are equal in status.

**Mutual benefits:**
Each party gains from federation

---

Definition of Federated Learning

- Interpretation of Federated Learning:
  - Models --- Sheep
  - Data --- Grass

- Originally, one need to purchase grass from different sources to feed sheep --- Companies gather lots of data to train models, where many challenges exist, such as user privacy, data security and regulations.

- Federated Learning provides an alternative: sheep are led to different farms and can thus eat grass from all places without having to move the grass. --- Federated learning models gather knowledge from various sources of data without having to observe them.
Key Components in Federated Learning

- Model design and hyperparameter tuning, e.g. number of layers, CNN or RNN, etc.
- Distributed learning algorithm (Chapter 3), e.g. client selection, tackling non-IID (or even contradictory) training data, system-algorithm co-design, etc.
- Communication optimization, e.g. alleviating the influence of network delay, model/gradient compression, etc.
- Security and privacy (Chapter 2), e.g. Homomorphic Encryption (HE), Differential Privacy (DP), Secure Multi-party Computation (MPC), etc.
- Incentive mechanism (Chapter 7), e.g. motivating organizations from different industries, adequate revenue allocation, etc.
Federated Learning Systems: Overview

Taxonomy of federated learning systems (FLSs)

Federated Learning System: Overview

- **Federated Optimization and Learning Algorithms**
  - HFL
  - VFL
  - FTL

- **Cross-site Communication**
- **Incentive Mechanism**
- **Preventing any info leakage**
  - HE
  - DP
  - MPC

- **Defense against model/data poisoning attacks**
  - FL has built-in mechanism for robustness, such as defending model/data poisoning attacks.
  - FL is more than what “MPC+ML” is about.

- **Preventing any info leakage**
  - The built-in MPC block in FL can prevent any info leakage (either model or data leakage).
  - FL can do whatever “MPC+ML” can do.
Federated Learning Applications

- Federated Learning + Other ML Algorithms (Chapter 8)
  - Federated learning + Computer vision (FL+CV)
  - Federated learning + Natural language processing (FL+NLP), including automatic speech recognition (ASR)
  - Federated learning + Recommender system (FL+RS)

- Federated Learning + Industry/Society, (Chapter 10 and cases from WeBank: https://www.fedai.org/cases/)
  - Federated learning + Finance, FinTech
  - Federated learning + Insurance, InsurTech
  - Federated learning + Healthcare
  - Federated learning + Education
  - Federated learning + AIoT
  - Federated learning + Smart City
  - Federated learning + Edge computing
  - Federated learning + 5G/6G

- Reference:
  - https://zhuanlan.zhihu.com/p/87777798
Federated Learning Datasets

- WeBank FedVision - Street Dataset, Available: [https://dataset.fedai.org/#/](https://dataset.fedai.org/#/)
  - A real-world **object detection** dataset that annotates images captured by a set of street cameras based on object present in them, including 7 classes. In this dataset, each or every few cameras serve as a device.

  - Federated Extended MNIST (FEMNIST ), 62 classes, Image Classification
  - Twitter, Sentiment140, Sentiment Analysis, federated
  - Shakespeare, Next-Character Prediction, federated
  - Celeba, Image Classification (Smiling vs. Not smiling), federated
  - Synthetic Dataset, Classification, federated
Federated Learning Open-source Platforms

- **WeBank FATE**, supports TensorFlow and PyTorch, https://github.com/FederatedAI/FATE
- Google TensorFlow Federated (TFF), https://github.com/tensorflow/federated
- OpenMined PySyft, supports PyTorch, https://github.com/OpenMined/PySyft
- PaddlePaddle/PaddleFL by Baidu, https://github.com/PaddlePaddle/PaddleFL
Privacy-Preserving Technologies

- Secure Multi-party Computation (MPC)
- Homomorphic Encryption (HE)
- Yao’s Garbled Circuit
- Secret sharing
- Differential Privacy (DP)

......
Secure Multi-Party Computation (MPC)

- Provides security proof in a well-defined simulation framework
- Guarantees complete zero knowledge
- Requires participants’ data to be secretly-shared among non-colluding servers
- **Drawbacks:**
  - Expensive communication,
  - Though it is possible to build a security model with MPC under lower security requirement in exchange for efficiency
Yao’s Garbled Circuit Protocol (Andrew Yao, 1986)

Garbled Circuit Protocol

Steps
• Alice builds a garbled circuit;
• Alice sends her input keys;
• Alice and Bob do Oblivious Transfer;
• Bob gets the output and sends back to Alice;
• Alice and Bob learn nothing about the other value.

Oblivious Transfer

Function f

Alice input a
Bob input b

Alice gets f(a,b) but learns nothing about Bob
Bob gets f(a,b) but learns nothing about Bob
SecureML: Privacy-preserving machine learning for linear regression, logistic regression and neural network training

- Combines secret sharing, garbled circuits and oblivious transfer
- Learns via two un-trusted, but non-colluding servers
- Computationally expensive

Differential Privacy

Definition: Differential Privacy (DP) [Dwork 2008]

A randomized mechanism \( M \) is \( \epsilon \)-differentially private, if for all output \( t \) of \( M \), and for all databases \( D_1 \) and \( D_2 \) which differ by at most one element, we have

\[
Pr(M(D_1) = t) = e^\epsilon Pr(M(D_2) = t).
\]

**Intuition:** changes in the distribution are too small to be perceived with variations on a single element.

Homomorphic Encryption

• Full Homomorphic Encryption and Partial Homomorphic Encryption.
• Paillier partially homomorphic encryption

\[
\text{Addition: } [u] + [v] = [u+v] \\
\text{Scalar multiplication: } n[u] = [nu]
\]

• For public key \( pk = n \), the encoded form of \( m \in \{0, \ldots, n-1\} \) is

\[
\text{Encode}(m) = r^n (1 + n)^m \mod n^2
\]

\( r \) is randomly selected from \( \{0, \ldots, n-1\} \).

• For float \( q = (s, e) \), encrypt \([q] = ([s], e)\), here \( q = s\beta^e \) is base-\( \beta \) exponential representation.

Applying HE to Machine Learning

Polynomial approximation for logarithm function

\[
\log \left( \frac{1}{1 + \exp(u)} \right) \approx \sum_{j=0}^{k} a_j u^j
\]

Encrypted computation for each term in the polynomial function

\[
\text{loss} = \log 2 - \frac{1}{2} y w^T x + \frac{1}{8} (w^T x)^2
\]

\[
[[\text{loss}]] = [[\log 2]] + \left( -\frac{1}{2} \right) * [[yw^T x]] + \frac{1}{8} [[(w^T x)^2]]
\]

- Kim, M.; Song, Y.; Wang, S.; Xia, Y.; and Jiang, X. 2018. Secure logistic regression based on homomorphic encryption: Design and evaluation. JMIR Med Inform 6(2)

- Y. Aono, T. Hayashi, T. P. Le, L. Wang, Scalable and secure logistic regression via homomorphic encryption, CODASPY16
Is the Gradient Info Safe to Share?

Algorithm ensures that no information is leaked to the semi-honest server, provided that the underlying additively homomorphic encryption scheme is secure*.

* Q. Yang, Y. Liu, T. Chen, Y. Tong, Federated machine learning: concepts and applications, ACM TIST, 2018

Horizontal Federated Learning, HFL

- Parties own data with overlapping features, i.e. aligned feature space; yet the training samples are different.
  - Also known as “cross-sample federated learning”, “feature-aligned federated learning”.
  - The feature space is identical.
  - HFL expands the number of training samples, with the feature dimensionality unchanged.

Horizontally partitioned data: data frames are partitioned horizontally into rows, each of which having the same features.

References:


Horizontal Federated Learning: Divide by Users/Samples

Step 1: Participants compute training gradients locally
- mask gradients with encryption, differential privacy, or secret sharing techniques
- all participants send their masked results to server

Step 2: The server performs secure aggregation without learning information about any participant

Step 3: The server sends back the aggregated results to participants

Step 4: Participants update their respective model with the decrypted gradients
Horizontal Federated Learning

- Multiple clients, one server
- Data is horizontally split across devices, homogeneous features
- Local training
- Selective clients


Vertical Federated Learning, VFL

- Parties hold data with identical data ID (i.e. training samples), but with different features.
  - A.k.a “Cross-feature federated learning”, “sample-aligned federated learning”. Suitable for federated learning across industries.
  - Before training, we take the intersection of data IDs held by different parties.
  - VFL increases data dimensionality at the cost of sample size (due to intersection of IDs).

Vertically partitioned data: partition data frames into columns, with each column holding the same feature.

Vertical Federated Learning

**Objective:**
- Party (A) and Party (B) co-build a FML model

**Assumptions:**
- Only one party has label Y
- Neither party wants to expose their X or Y

**Challenges:**
- Parties with only X cannot build models
- Parties cannot exchange raw data by law

**Expectations:**
- Data privacy for both parties
- Model is LOSSLESS

---

<table>
<thead>
<tr>
<th>ID</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>9</td>
<td>80</td>
<td>600</td>
</tr>
<tr>
<td>U2</td>
<td>4</td>
<td>50</td>
<td>550</td>
</tr>
<tr>
<td>U3</td>
<td>2</td>
<td>35</td>
<td>520</td>
</tr>
<tr>
<td>U4</td>
<td>10</td>
<td>100</td>
<td>600</td>
</tr>
<tr>
<td>U5</td>
<td>5</td>
<td>75</td>
<td>600</td>
</tr>
<tr>
<td>U6</td>
<td>5</td>
<td>75</td>
<td>520</td>
</tr>
<tr>
<td>U7</td>
<td>8</td>
<td>80</td>
<td>600</td>
</tr>
</tbody>
</table>

**Retail A Data**

<table>
<thead>
<tr>
<th>ID</th>
<th>X4</th>
<th>X5</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>6000</td>
<td>600</td>
<td>No</td>
</tr>
<tr>
<td>U2</td>
<td>5500</td>
<td>500</td>
<td>Yes</td>
</tr>
<tr>
<td>U3</td>
<td>7200</td>
<td>500</td>
<td>Yes</td>
</tr>
<tr>
<td>U4</td>
<td>6000</td>
<td>600</td>
<td>No</td>
</tr>
<tr>
<td>U8</td>
<td>6000</td>
<td>600</td>
<td>No</td>
</tr>
<tr>
<td>U9</td>
<td>4520</td>
<td>500</td>
<td>Yes</td>
</tr>
<tr>
<td>U10</td>
<td>6000</td>
<td>600</td>
<td>No</td>
</tr>
</tbody>
</table>

**Bank B Data**
Privacy-Preserving Entity Match

- Party A and B learns their overlapping IDs but nothing else
Vertical Federated Learning

Federated model

Encrypted model training

Encrypted entity alignment

Corp. A
No data exchange

Corp. B

Model A

Model B

Data from A

Data from B

Collaborator C

1. Sending public keys
2. Exchanging intermediate results
3. Computing gradients and loss
4. Updating models

Privacy-Preserving inference

• Suppose a new user ID arrives at Party B,

Step 1
Party B sends encrypted ID to A

Step 2
Party A and B compute local results

Step 3
Party A sends its results to B

Step 4
Party B combines results

Find ID?
Yes
No

Party A

Party B

Encrypted ID matching

sends sub-model score $U_A$

Add sub-model scores $U_A + U_B$

Sends end signal

\[ \text{Privacy-Preserving inference} \]

\[ \text{• Suppose a new user ID arrives at Party B,} \]

\[ \text{Step 1} \]

\[ \text{Party B sends encrypted ID to A} \]

\[ \text{Step 2} \]

\[ \text{Party A and B compute local results} \]

\[ \text{Step 3} \]

\[ \text{Party A sends its results to B} \]

\[ \text{Step 4} \]

\[ \text{Party B combines results} \]

\[ \text{Find ID?} \]

\[ \text{Yes} \]

\[ \text{No} \]

\[ \text{sends sub-model score } U_A \]

\[ \text{Add sub-model scores } U_A + U_B \]

\[ \text{Sends end signal} \]
Security Analysis

• Security against third-party C
  - all C learns are the masked gradients and the randomness and secrecy of the masked matrix are guaranteed

• Security against each other
  - Party A learns its gradient at each step, but this is not enough for A to learn any information from B
    - inability of solving n equations in more than n unknowns

• Security in the semi-honest setting
XGBoost in Federated Learning

SecureBoost

Figure 1: Illustration of the proposed SecureBoost framework

Kawei Cheng, Tao Fan, Yilun Jin, Yang Liu, Tianjian Chen, Qiang Yang, SecureBoost: A Lossless Federated Learning Framework, IEEE Intelligent Systems 2020

GBDT in HFL

Qinbin Li, Zeyi Wen, Bingsheng He, Practical Federated Gradient Boosting Decision Trees, AAAI, 2019
Federated Transfer Learning, FTL

- Parties hold data with different ID and different features (some parties may not have labels).
  - Suitable for federated learning across industries.
  - Common methods include model transfer, instance transfer, feature transfer and domain adaptation, etc.
  - Peer-to-peer architecture is commonly used.
  - FTL is important to solving the problem of ‘small data’ and ‘unlabeled data’.

Federated Transfer Learning

Source Domain Party A

Target Domain Party B

Step 1
Party A and B send public keys to each other

Step 2
Parties compute, encrypt and exchange intermediate results

Step 3
Parties compute encrypted gradients, add masks and send to each other

Step 4
Parties decrypt gradients and exchange, unmask and update model locally

Efficiency: BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning

- Reducing the encryption overhead and data transfer
  - Quantizing a gradient value into low-bit integer representations
  - Batch encryption: encoding a batch of quantized values to a long integer

- BatchCrypt is implemented in FATE and is evaluated using popular deep learning models
  - Accelerating the training by 23x-93x
  - Reducing the network footprint by 66x-101x
  - Almost no accuracy loss (<1%)

C. Zhang, S. Li, J. Xia, W Wang, F Yan, Y. Liu, BatchCrypt: Efficient Homomorphic Encryption for Cross-Silo Federated Learning, USENIX ATC’20 (accepted)
Incentivize Parties to Join: Federated Learning Exchange

- **Observation:** The success of a federation depends on data owners to share data with the federation.
- **Challenge:** How to motivate continued participation by data owners in a federation?

**Research Question:** How to determine $u_i(t)$?

# Federated Learning vs Federated Database vs Blockchain

<table>
<thead>
<tr>
<th>Consistency (Unique Results)</th>
<th>Federated Learning</th>
<th>Federated Database</th>
<th>Blockchain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training and Inference Results</td>
<td>Query Results</td>
<td>Transaction Information</td>
</tr>
<tr>
<td>Atomicity (Same Status)</td>
<td>True for Models</td>
<td>Partially True</td>
<td>True</td>
</tr>
<tr>
<td>Collective Security (Multi-party Participation, Secure Computation)</td>
<td>True</td>
<td>False</td>
<td>N/A</td>
</tr>
</tbody>
</table>

- **Summary**
  - **Federated Learning**: No data replication, Secure model parameter transfer, Data and model security guarantee.
  - **Federated Database**: Interoperability of multiple databases, No security requirement, Possible security leak in the central management system
  - **Blockchain**: Distributed ledger, Data replicated on all nodes
Advances and Open Problems in Federated Learning

Peter Kairouz7* H. Brendan McMahan7* Brendan Avent21 Aurélien Bellet9 Mehdi Bennis19 Arjun Nitin Bhagoji13 Keith Bonawitz7 Zachary Charles7


Ben Hutchinson7 Justin Hsu25 Martin Jaggi4 Tara Javid17 Gauri Joshi2 Mikhail Khodak2 Jakub Konečný7 Aleksandra Korolova21 Farinaz Koushanfar17 Sanmi Koyejo7,18 Tancrède Lepoint7 Yang Liu12 Prateek Mittal13 Mehryar Mohri7 Richard Nock1 Ayfer Özgür15 Rasmus Pagh7,10

Marina A. Rostami11 Daniel Ramage7 Ramesh Raskar11 Dawn Song16 Weikang Song7 Sebastian U. Stich4 Ziteng Sun3

Ananda Theertha Suresh7 Florian Tramèr15 Praneeth Vepakomma11 Jianyu Wang2 Li Xiong5 Zheng Xu7 Qiang Yang8 Felix X. Yu7 Han Yu12 Sen Zhao7

1Australian National University, 2Carnegie Mellon University, 3Cornell University, 4École Polytechnique Fédérale de Lausanne, 5Emory University, 6Georgia Institute of Technology, 7Google Research, 8Hong Kong University of Science and Technology, 9INRIA, 10IT University of Copenhagen, 11Massachusetts Institute of Technology, 12Nanyang Technological University, 13Princeton University, 14Rutgers University, 15Stanford University, 16University of California Berkeley, 17University of California San Diego, 18University of Illinois Urbana-Champaign, 19University of Oulu, 20University of Pittsburgh, 21University of Southern California, 22University of Virginia, 23University of Warwick, 24University of Washington, 25University of Wisconsin–Madison
Applications

- Anti money laundering
  - Recall improves 15%
  - Audit efficiency improved by over 50%

- Internet + banking
  - Risk modeling
  - Performance keeps increasing with respect to the enriched features

- Internet + insurance
  - Insurance pricing
  - Pricing model improves accuracy
  - Coverage ratio is over 90%

- Internet + retailers
  - Intelligent marketing
  - Marketing efficiency improves greatly;
  - Better user profile and targeting;
# UC2: Risk Management with Federated Learning

## Other Enterprise

<table>
<thead>
<tr>
<th>ID (纳税人识别号)</th>
<th>X1 (近3个月开票金额)</th>
<th>X2 (近6个月开票金额)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>90</td>
<td>150</td>
</tr>
<tr>
<td>U2</td>
<td>40</td>
<td>100</td>
</tr>
<tr>
<td>U3</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>U4</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>U9</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>U10</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>U11</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

## WeBank

<table>
<thead>
<tr>
<th>ID (纳税人识别号)</th>
<th>X3 (央行征信分)</th>
<th>Y (逾期概率)</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>600</td>
<td>0.79</td>
</tr>
<tr>
<td>U2</td>
<td>550</td>
<td>0.11</td>
</tr>
<tr>
<td>U3</td>
<td>520</td>
<td>0.88</td>
</tr>
<tr>
<td>U4</td>
<td>600</td>
<td>0.15</td>
</tr>
<tr>
<td>U5</td>
<td>600</td>
<td>0.90</td>
</tr>
<tr>
<td>U6</td>
<td>520</td>
<td>0.81</td>
</tr>
<tr>
<td>U7</td>
<td>600</td>
<td>0.32</td>
</tr>
</tbody>
</table>

## Data Dimensions

- **Other Enterprise**: 200 Dimensions
- **WeBank**: 20 Dimensions

## Notes

- 60 Million
- 400 K
- 200 Dimensions
- 20 dim
- tax, bank, business
Construction-Site Safety w/ Federated Computer Vision

W ebank AI X Extreme Vision
Federated Recommendation

Assumption: a trustworthy 3rd-party as coordinator, which can be removed
Horizontal Federated Recommendation

Example: movie recommendation with data from individual users
Federated Collaborative Filtering [Ammad et al. 2019]

Intuition: decentralized matrix factorization, each user profile is updated locally, item profiles are aggregated and updated by coordinator.
IEEE International Standard for Federated Learning

https://sagroups.ieee.org/3652-1/

Become a Member:
WeBank AI: Federated Learning Open Source

1月 微众AI亮相 AAAI 会议 发布联邦学习开源框架 FATE0.1版本

2月 微众AI领衔推动联邦学习国际标准制定

3月 微众AI主导首个国际联邦学习学术研讨会

4月 微众AI主要合作伙伴切换到FATE 微众与瑞士再保险签订战略合作协议，推动联邦学习在再保险业应用

5月 微众AI牵头的国内首个联邦学习标准正式出台

6月 微众联邦学习开源项目加入Linux基金会
Challenges for Federated Learning

**Models [BVH+18]**


**Data [HAP17]**

Vertical Federated Learning

**Banking Authentication :**
1. Upload front camera photo: to judge whether an image is taken from a real person or his photo.
2. Bank A can cooperate with a **Secure Face Database** with VFL;

**What can AutoFL (Vertical) do :**

To determine the learning architecture **automatically** and **locally** in a **communication-efficient manner** with **data protection**;
Privacy Attack Example: Deep Leakage.

Song Han from MIT designed Deep Leakage Attacks that tackle DP-protected models, and are able to reconstruct training data from gradients with pixel-level accuracy.

References