Federated AI in Future Digital Banking

Tianjian Chen
Deputy GM of AI Dept.
Aug. 2019
Agenda

01  WeBank and Digital Banking in China
02  Basics and Applications of Federated Learning
03  Introduction to the FATE Project
01
WeBank and Digital Banking
China’s 1st Digital Bank
Open in 2014

Tech Staff
> 50%

Operations
7*24

Time-to-market
Product Record
11 DAY

> 150 Million Customer

> 10B Annual Revenue

30% Tencent

Tencent Open in 2014

> > 10B Annual Revenue

> > 150 Million Customer

> > 50% Tech Staff

> 7*24 Operations

> 11 DAY Time-to-market Product Record
Adventure of Globalization

- 2001: China Join WTO
- 2008: Financial Crisis
- 1990: Gulf War, NATO participates
- 1993: Bosnian War, Russia participates
- 2000: India-Pakistan, Russia- Chechnia Wars
- 2004: Impact of Iraq & Afghanistan Wars
- 2007-2009: Military expansion by NATO countries & Russia
- Emerging market economies take over during financial crisis & eurozone crisis
- The globalized world is turning multipolar

Src: Credit Swiss
2009
Massive Construction of 3G/4G Network & High Speed Rail
Creation of An Unprecedented Consumer Market On Mobile

Consumption in China and the US

- Chinese household consumption ($bn)
- US personal consumption ($bn)
- Ratio (US/China) Forecast

In 2000, US consumption levels were 13 times larger than China's...

...but are now only three times larger—and the gap is closing

The Amount of China's Online Banking Transactions 2011-2019

Source: The data were calculated based on the financial results published by enterprises and interviews with experts in iResearch statistical model.

©May 2016 iResearch Global Inc. www.iresearchchina.com
Banking Goes Full Digital by AI Technology

- Online Acquisition
- Realtime-Bidding
- Transfer Learning
- 15x Cost-Efficiency
Customer Inquiries

- 4 million Sessions / day
- Reinforcement Learning
- 10x Cost-Efficiency
Identity Verification

- 1 million query / day
- Deep Learning
- 10x Cost-Efficiency
Risk Evaluation

- 200+ Models / 200k+ Variables
- Large Scale Machine Learning
- 3x Cost-Efficiency
AI Driven Experience

- Get Loan in 60 secs
- Anywhere / Anytime
- No Extra Fee Charged
STORM IS COMING
Data Protection-ism：Inevitable
Data Regulations in China

- **Data Regulations in China**
  - **National Law:** Various legal documents and regulations.
  - **Administrative Regulation:** Administrative regulations related to data security.
  - **Departmental Regulations:** Regulations issued by relevant departments.

**Timeline:**
- **2009.01.28:** The decision of the Standing Committee of the National People's Congress on Strengthening the Protection of Network Information.
- **2012.12.28:** Amendments to the Criminal Law.
- **2015.08.29:** Scientific Data Management Regulations.
- **2016.11.07:** The Cybersecurity Law of the People's Republic of China.
- **2018.03.17:** Opinions on Data Security Management (Draft征求意见稿).
- **2018.07.12:** The National Health Medical Big Data Standards, Security, and Service Management Regulations (Trial). (全面化: 各领域数据管理细则密集出台，用户授权+监管部门审批)
- **2018.08.31:** The Commercial Code of the People's Republic of China.
- **2018.08.31:** The Amendment to the Criminal Law (Nine).
- **2019.05.28:** The Cybersecurity Law of the People's Republic of China.

**Key Points:**
- **Detail:** Various legal and regulatory frameworks have been established in China to protect data security and privacy.
- **Enforcement:** There is a clear accountability for data controllers, and penalties are applied to natural persons.
Federated Learning Comes to Rescue

02 Basics and Applications of Federated Learning
Google Federated Learning Concept 2017

H. Brendan McMahan et al

Communication-Efficient Learning of Deep Networks from Decentralized Data

Google, 2017


Practical secure aggregation for privacy-preserving machine learning

Google, 2017
General Federated Learning Catalog

Vertical Federated Learning

Horizontal Federated Learning

Federated Transfer Learning

Video:
https://youtu.be/0Ai7vtymJzl
S0. Sample Alignment based on Private Set Intersection

**Party A**

\(X_A: \{u_1, u_2, u_3, u_4\}\)

\(Y_A = \{ri^e \cdot H(u_i) | u_i \in X_A, ri: \text{rand}\}\)

\(D_A = \{H(ri \cdot H(u_i)^d)^d / ri) = H((H(u_i))^d) | ri \cdot H(u_i) \in Z_A\}\)

\(Z_A = \{H((H(u))^d)^d) = ri \cdot (H(u))^d \mod n | ri \cdot H(u) \in Y_A\}\)

\(I = D_A \cap Z_B\)

\(l = \{H((H(u))^d) | u \in X_B\}\)

\(I, D_A \Rightarrow \{u_1, u_2, u_3\}\)

**Party B**

\(X_B: \{u_1, u_2, u_3, u_5\}\)

**public key:** \((n, e)\)

\(Z_B = \{H((H(u))^d) | u \in X_B\}\)

\(I, Z_B \Rightarrow \{u_1, u_2, u_3\}\)

**RSA:** \(n, e, d\)

**H(x): Hashing on X**
S1. Federated Model Training Based on HE

Polynomial approximation for logarithm function

\[ I(w) = \log(1 + \exp(-yw^T x)) \]
\[ \approx \log 2 - \frac{1}{2} yw^T x + \frac{1}{8} (w^T x)^2 \]

\[ \nabla I(w) = \left( \frac{1}{1 + \exp(-yw^T x)} - 1 \right) yx \]
\[ \approx \left( \frac{1}{2} yw^T x - 1 \right) \frac{1}{2} yx \]

Encrypted computation for each term in the polynomial function

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

- 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
UseCase：Small Business Loan Risk Estimation

Ideal Big Data for Risk Estimation

Skinny Reality

对客户缺乏全面了解
• 通常只有央行信用报告
• 有重复表现的客户小于10%
• 70%客户无任何信用表现

全面的企业及其控制人相关数据
UseCase: Risk Estimation based on Federated Learning

通过发票数据、央行征信分等标签属性进行联合建模，预测小微企业信贷逾期概率

<table>
<thead>
<tr>
<th>ID</th>
<th>X1 近3个月发票金额</th>
<th>X2 近6个月发票金额</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>

<table>
<thead>
<tr>
<th>ID</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>

发票数据衍生变量

发票数据衍生变量：

- 发票金额
- 纳税人识别号

税务属性

200维度

金融属性

业务表现

20维度
Project Outcome

Model AUC +12%

M6 -40%
Other Applications

Re-Insurance Network

Insurance Network

Anti-Money Laundry Network

Retail Operation AI Network

AI Industry Network

Internet data

纵向联邦

横向联邦

保险公司1

保险公司2

银行1

银行2

视觉公司1

视觉公司2

客户1

客户2

客户3

客户4

WeBank

10
Federated AI Technology Enabler
Meet FATE : Federated AI Technology Enabler

Vision

• Industry Level Federated Learning Framework with Out-of-box Usability
• Enable Big Data Collaboration with Data Protection Regulation Compliance

Github

• https://github.com/WeBankFinTech/FATE

Website

• https://FedAI.org
Roadmap 2019

2019.02
FATE v0.1
Federated ML

201905
FATE v0.2
FATE-Serving

201908
FATE v1.0
FATE-FLOW | FATE-Board

201911
FATE v1.2
FATE-Script with GPU Support

201903
GitHub Star above 100
First Contributor Outside Webank

201906
FATE v0.3
Donated to Linux Foundation

201909
FATE v1.1
Secret Sharing
Deep Learning HP
Federated Analytics

201912
FATE v1.3
Encrypted Deep Model
Project Landscape

FATE-Flow
- Flow DAG Parser
- Lifecycle Manager
- Multi-Party Task Scheduler

FATE-Serving
- Federated Inference
- Model Manager
- Version Control

Fate-Board
- FL Visualization
- Monitoring
- Log Manager

FederatedML: Federated Machine Learning Core Component

EggRoll: Distributed Computing Framework

Data Injection
- Access Interface
  - HIVE
  - MySQL
  - Level DB
- Format Adapter
  - Amazon S3
  - CSV
  - HBASE
  - HDFS
  - ......

Federated Network: Cross-Site Networking

Device
- CPU Clusters
- GPU Clusters
- Android / IOS
## FATE FederatedML

**Federated Learning Algorithm Library**

### Algorithms
- Secure Intersection
- Secure Federated Feature Engineering
- Secure LR
- Secure Boost
- Secure DNN/CNN
- Secure FTL

### ML Operator
- Federated Aggregator
- Activation
- Regulation
- Loss
- Optimizer
- Gradient
- Hessian

### Numeric Operator
- Add
- Sub
- MUL
- DIV
- Comparison
- AND
- OR
- Scalar Product

### MPC Protocol
- Homomorphic Encryption
- Secret-Sharing
- Oblivious Transfer
- Garbled Circuit
- RSA
Federated Statistics → Federated Feature Engineering → Federated Training → Federated Evaluation → Model Management → Model Publish

Train & evaluate data

PSI → Federated Feature Binning → Federated Feature Selection

FATE-Serving FATE-Serving FATE-Serving

WeBank
**FATE Flow**

Federated Training Process Coordinator

- **Proxy Service**
  - Federated Task Scheduler
  - Some Other Party

- **Server**
  - FATE-Flow Client
  - FATE-Board
  - FATE-Flow Service
  - Job Queue
  - Job Scheduler
  - DSL Parser
  - Job Controller
  - Model Manager
  - Federation Service
  - FederatedML
  - Storage & Computing

- **Metadata Service**
  - DB

- **FATE-Flow Clients**
  - FATE-Flow Client
  - FATE-Board
FATE-Serving
Federated Inference & Version Control

- Inference & Management API
- Dynamic Loaders
- Federated Inference Service
  - Online FE
  - Inference Pipeline
  - GRPC
  - Batch Request
- Multi-Level Cache
  - Local cache
  - Distributed cache
  - Warming-up Schedule
- Snapshot Manager
- Storage

WeBank
Model Version Control System

- **Publish Load Request**
  - ServingGA1
  - ServingGB1
  - Dynamic loaders
  - Model object
  - Model pool
  - Model binding map
    - id0&v0:(my_id, my_v)
    - id1&v1:(my_id, my_v)
    - ...

- **Model Info**
  - ServingHA1
  - ServingHB1
  - Dynamic loaders
  - Model object
  - Model pool
  - Model binding map
    - id0&v0:(my_id, my_v)
    - id1&v1:(my_id, my_v)
    - ...

- **Publish Online AB Test**
  - ServingGA1
  - Model pool
  - Model default map
  - Guest
  - Model info
  - App factory
  - Model pool
  - Dynamic loaders
    - Input object
    - App object
    - Reflection
    - Model object

- **Host**
  - Dynamic loaders
  - Model binding map
  - Model object

- **Guest**
  - Dynamic loaders
  - Model binding map
  - Model object

- **Notes**
  - as party_id default model
  - as model_id default version
Typical Pipeline of Online Inference

req → Service
  ↓
  model selection
  ↓
  Specified
  ↓
  party id default
  ↓
  model id default
  ↓
  Loader from pool
  ↓
  processing
  ↓
  preprocessing-app
  ↓
  Inference Pipeline
  ↓
  Online FederatedML
  ↓
  Online FederatedML
  ↓
  Online FederatedML
  ↓
  processing
  ↓
  postprocessing-app
  ↓
  Remote Federated Result Cache

Federated Network

Service
  ↓
  model selection
  ↓
  Specified
  ↓
  Loader from pool
  ↓
  Inference Pipeline
  ↓
  Online FederatedML
  ↓
  Online FederatedML
  ↓
  Online FederatedML
FATE Board
Visualization and Management

Web UI

JOB DASHBOARD
- Dataset Info
- Job Graph
- Job Status
- Log

JOB VISUALISATION
- Job Summary
- Model Outputs
- Components Parameters
- Data Preview
- Job Graph
- Log Panel

RUNTIME MANAGEMENT
- Overview
- Life Cycle

CONFIGURATION MANAGEMENT
- Overview
- Role Management
FL Visualization  (Video)

FATEBoard

https://youtu.be/ksCJ9f4P1Rk
04
Call for Collaborations
Collaborations Under LF

- Cloud Native Computing Foundation
- Harbor
- KubeFlow
- Confidential Computing Consortium
- LF AI
- Angel
- Pyro
• An open source trusted cloud native registry.
• The first CNCF project originated from China.
• Thousands of production users worldwide.

**GitHub Stars** 8800+
**Contributors** 140+
**Commits** 7200+
**Forks** 2400+

**1000+ Production Users**

**20+ Partners**

**140+ Contributors**
Stars, Fork and Watch

https://github.com/Angel-ML

Key Stats

- 7 Sub Projects
  - math2/format/mlcore
  - Angel
  - Serving
  - SONA (Spark ON Angel)
  - PyTorch On Angel
- 250K+ lines of java code
- 50K+ lines of scala code
- 370+ Issues:
  - Open: 84
  - Closed: 290
- 360+ Pull Requests
  - Open: 8
  - Closed: 340
- 2000+ Commits
- 8 Committers
- 34 Contributors

1000+ new stars since join LF AI
Federated AI Ecosystem

Collaborative Learning and Knowledge Transfer with Data Protection

GitHub  White Paper  Subscribe Us

https://FedAI.org
Contact Us

- Website: https://FedAI.org/
- GitHub: https://github.com/WeBankFinTech/FATE
- YouTube Channel: https://www.youtube.com/c/FederatedAI
- Reference Materials: https://www.fedai.org/achiList
- Email: tobychen@webank.com
谢谢
Thanks