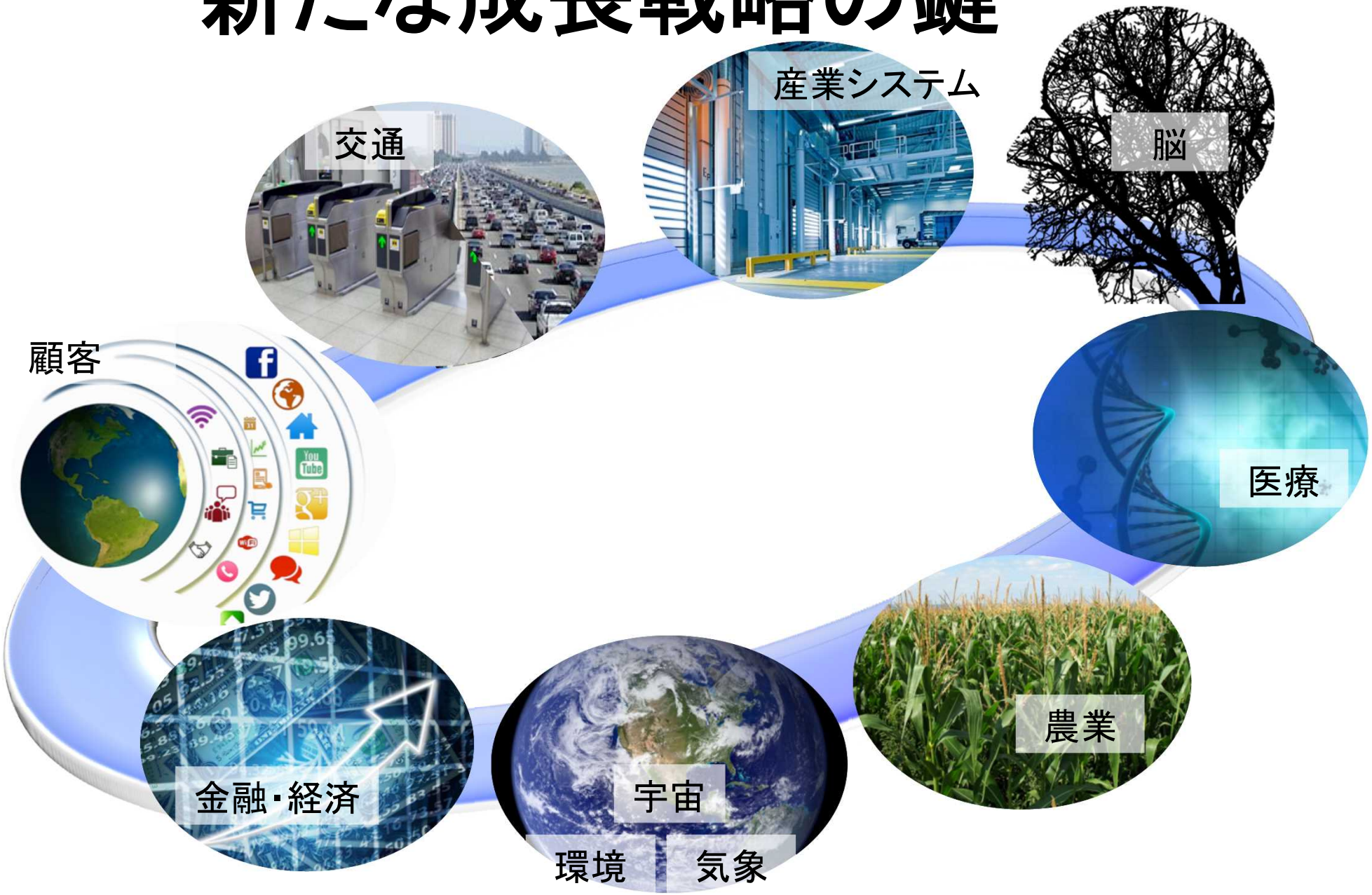
暗号技術とAIの融合による秘密計算技術

国立研究開発法人情報通信研究機構
サイバーセキュリティ研究所
セキュリティ基盤研究室
主任研究員　　レ　チュウ　フォン

2019年7月18日

データ統合利活用：
新たな成長戦略の鍵



データセキュリティの確保とプライバシー保護が課題

暗号技術＋人工知能技術

プライバシーを保護した状態でデータ分析

データ漏洩対策は大丈夫？

お客様のプライバシー確保

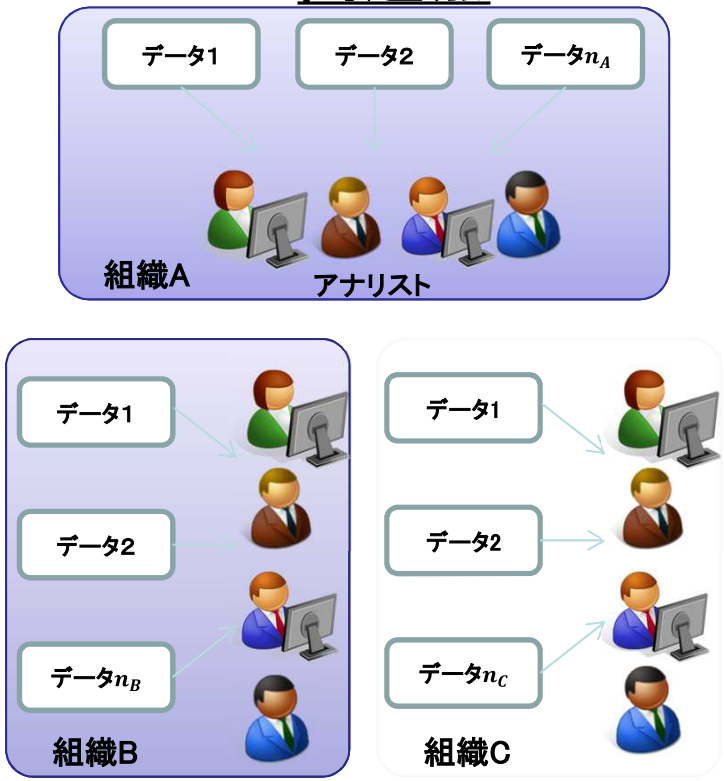
暗号により
データの機密性を
確保することで
組織横断でのデータ利活用を促進

次世代AI 技術によるデータ解析

新たな知見・イノベーション
多様な経済分野でのビジネス創出

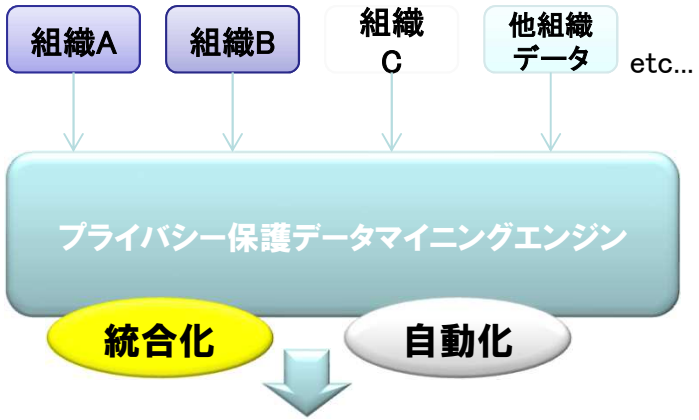
現状 と めざす構想

現状(従来技術)とその問題点



個々の**組織内**で分析、組織横断でのデータ利活用ができていない

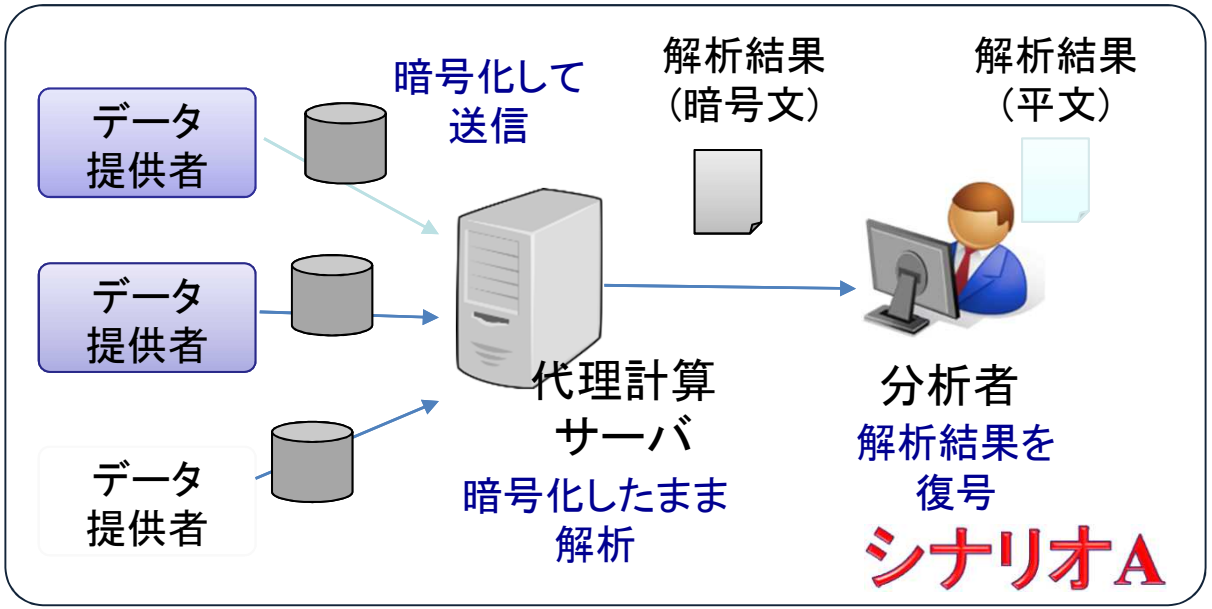
めざす構想

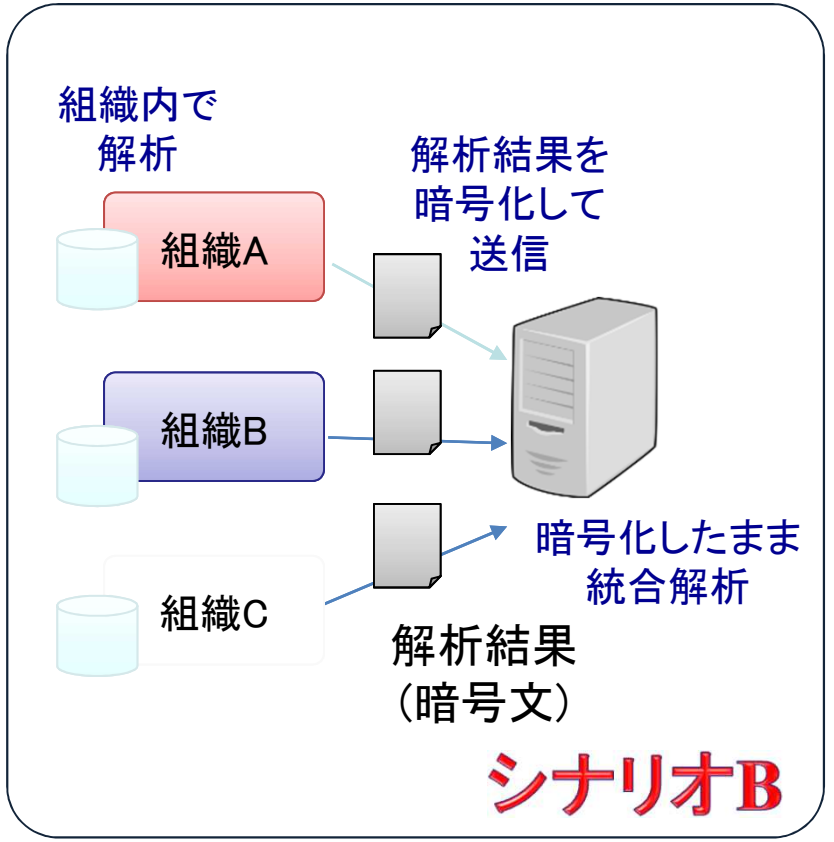


分析結果

- データを開示せずに、良い分析結果が得られる
- 応用:
  銀行の顧客のデータ、病院の患者のデータなど

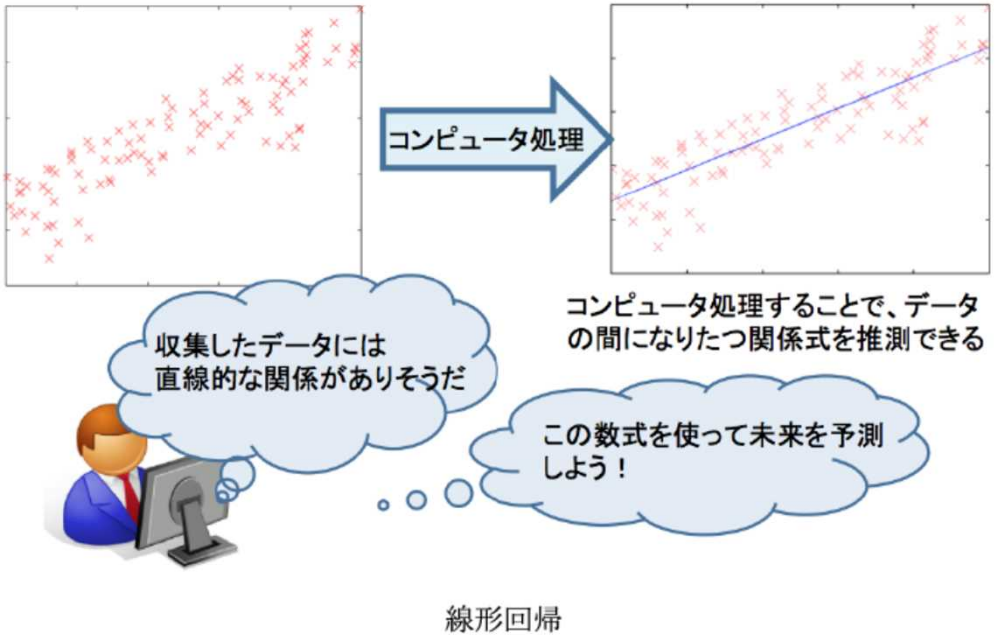
プライバシー保護データ解析利用シナリオ（二つ）





暗号化したままビッグデータ予測

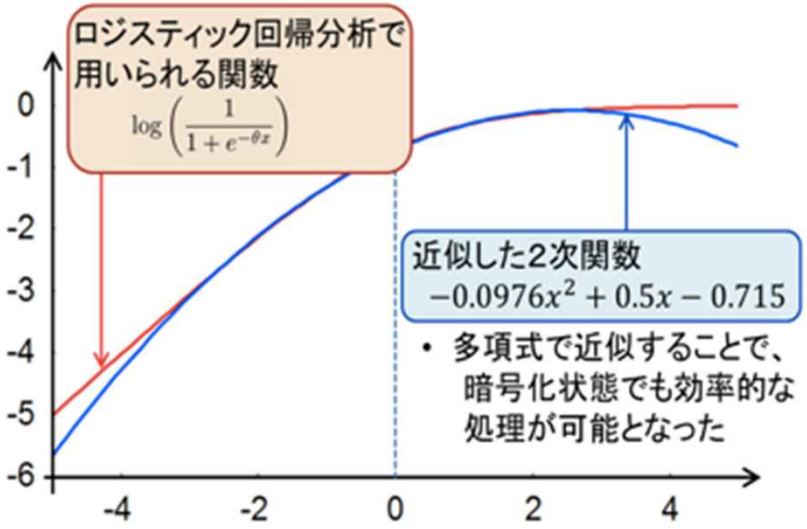
- ビッグデータ解析で多用されている線形回帰分析をデータを暗号化したまま計算可能に
- 100万件のデータに対する線形回帰計算を暗号化したまま行い、30分程度で処理ができることをシミュレーションで確認
- NICTプレスリリース(2015.1.19)：
  「暗号化状態でセキュリティレベルの更新と演算の両方ができる準同型暗号方式を開発」

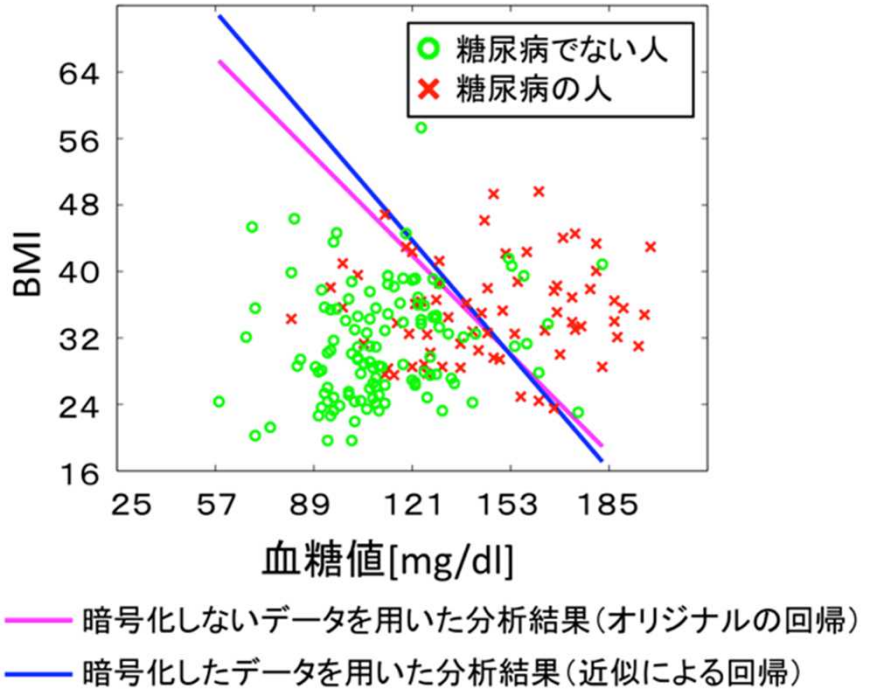


線形回帰

暗号化したままビッグデータ分類

- ビッグデータ解析で多用されているロジスティック回帰分析をデータを暗号化したまま計算可能に
- 暗号化された1億件のデータを30分以内で複数グループに　分類できることをシミュレーションで確認
  - NICTプレスリリース「**暗号化したままデータを分類できるビッグデータ向け解析技術を開発**」**(2016.1.14)**



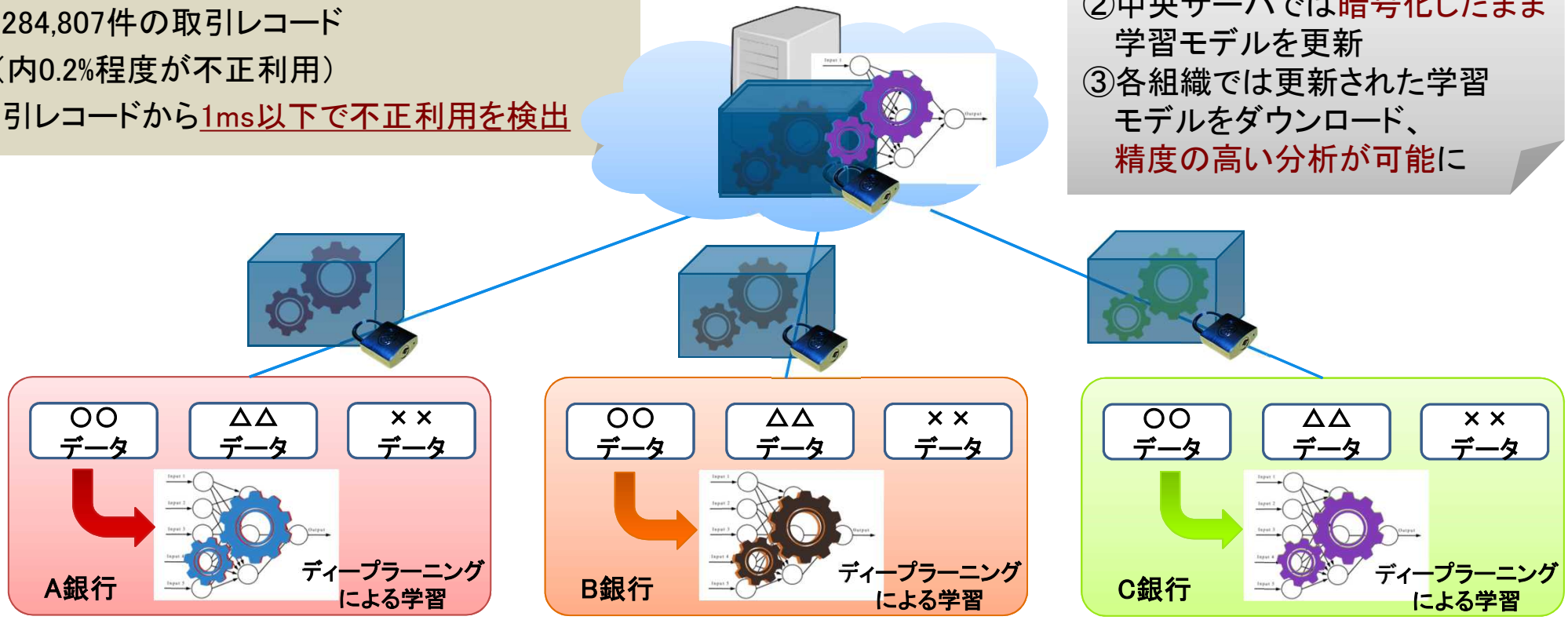


プライバシー保護 ディープラーニング（予測・分類）

- 組織が持つデータを外部に開示することなく深層学習を行うプライバシー保護深層学習システム

オープンデータセットを用いた実用性検証
◆欧州のクレジットカード取引データ
◆284,807件の取引レコード
（内0.2%程度が不正利用）
取引レコードから1ms以下で不正利用を検出

①各組織から学習済モデルのパラメータを暗号化して中央サーバに送信
②中央サーバでは暗号化したまま学習モデルを更新
③各組織では更新された学習モデルをダウンロード、精度の高い分析が可能に



複数組織で連携した分散協調型の深層学習

オープンデータを用いたプロトタイプの検証

オープンデータセットを用いた実用性検証
◆欧州のクレジットカード取引データ
◆284,807件の取引レコード
（内0.2%程度が不正利用）
取引レコードから1ms以下で不正利用を検出

複数組織からのデータ提供をシミュレーションするため、
公開データを計21個のグループに分割

- **学習用20セット＋テスト用1セット**

	不正利用	正当な利用
学習用データセット1,…,20	13 ~ 27	≈ 11379
テストセット	98	56863

プロトタイプ

ラウンド1

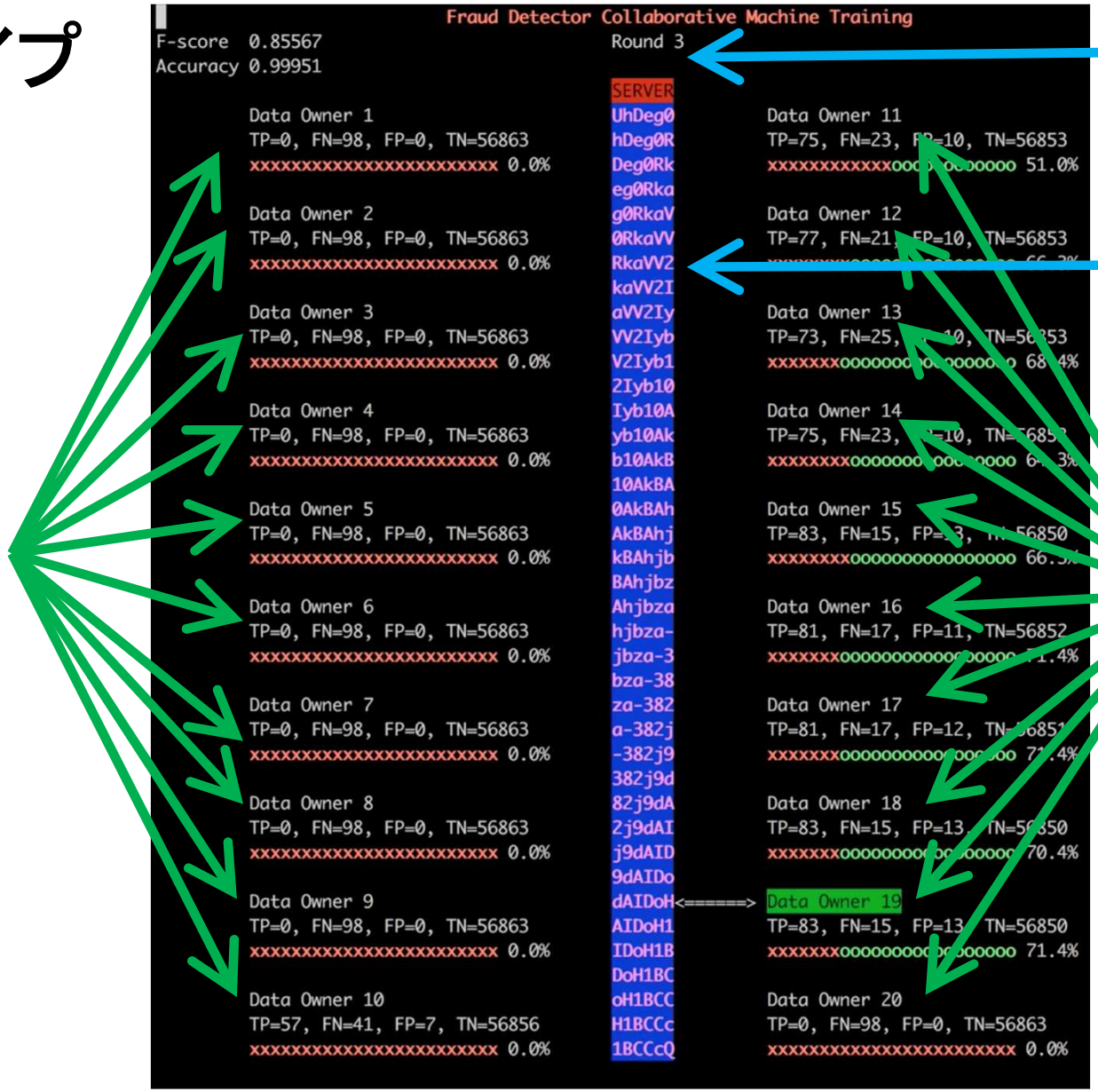
中央サーバ

データ
所有者
（計20組織）

データ
所有者
（計20組織）

学習ラウンド１終了時には、テストセット内の98件の不正利用が検出できていない

プロトタイプ



ラウンド3

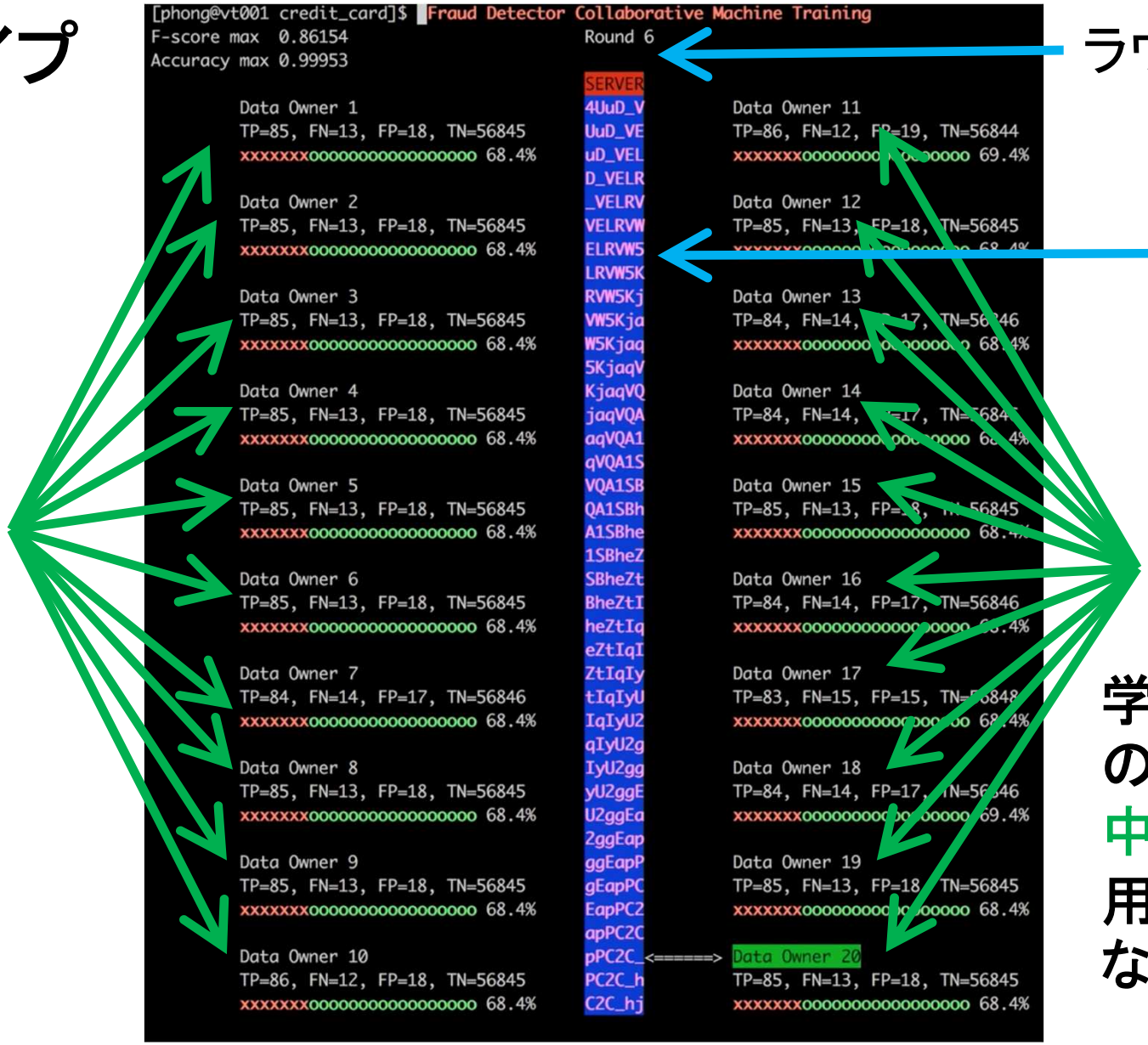
中央サーバ

データ
所有者
（計20組織）

データ
所有者
（計20組織）

学習が進むにつれて、**不正利用の検出率**が上がっていく

プロトタイプ



ラウンド6

中央サーバ

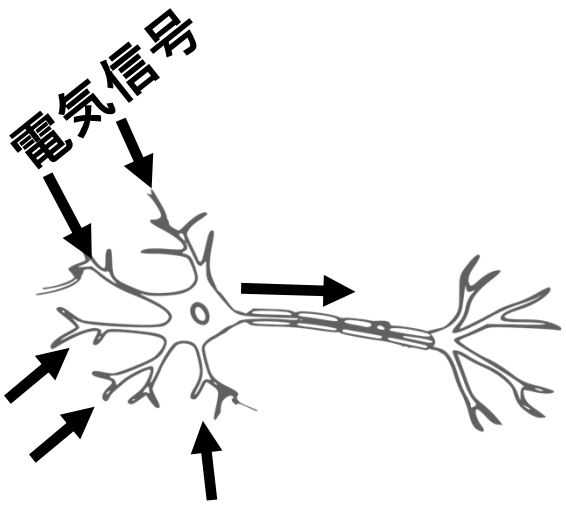
データ
所有者
（計20組織）

データ
所有者
（計20組織）

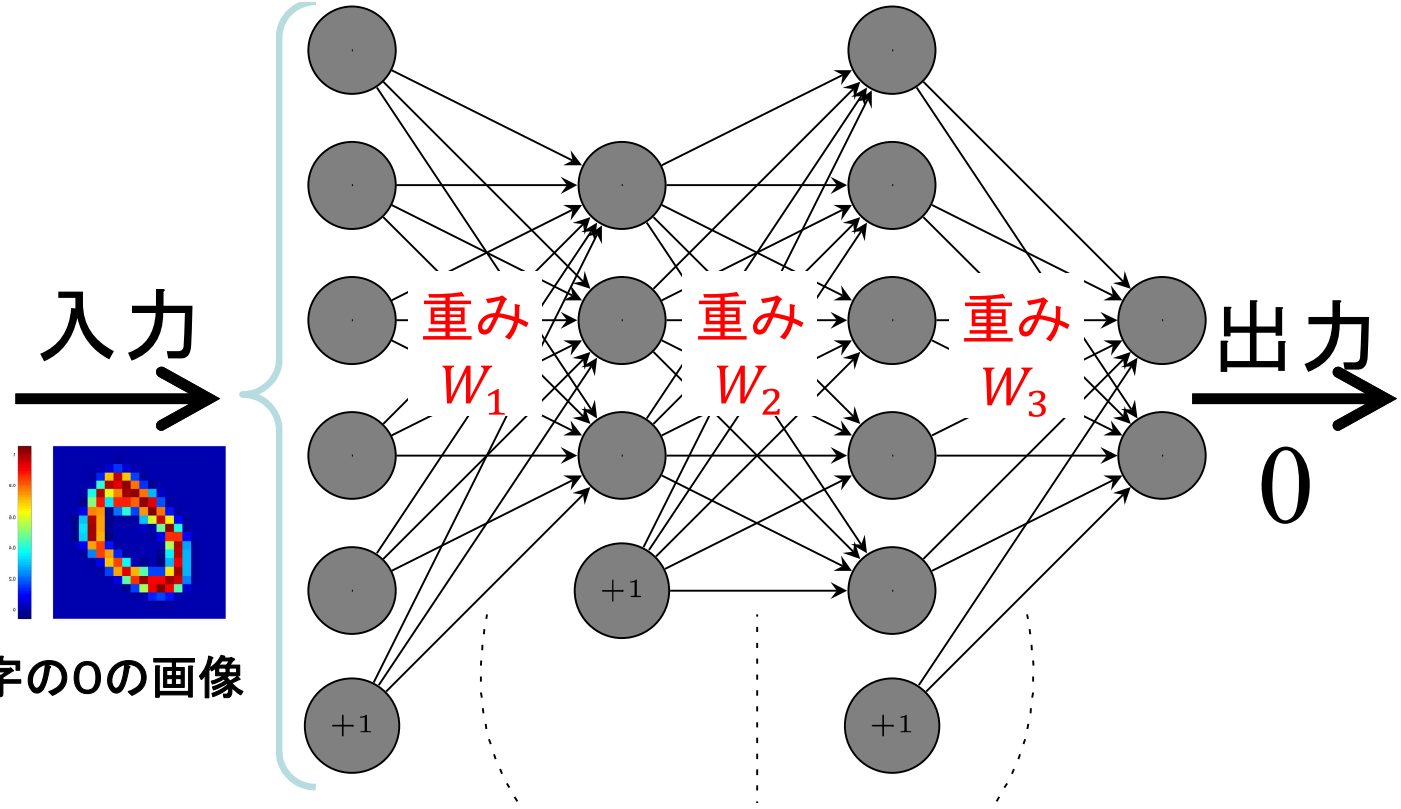
学習アルゴリズムの収束後、98件中85件の不正利用が検出可能になっている

提案システムのため
準備1/2：ニューラルネットワーク

生物の神経細胞
（ニューロン）



人工知能（AI）
のニューラルネットワーク

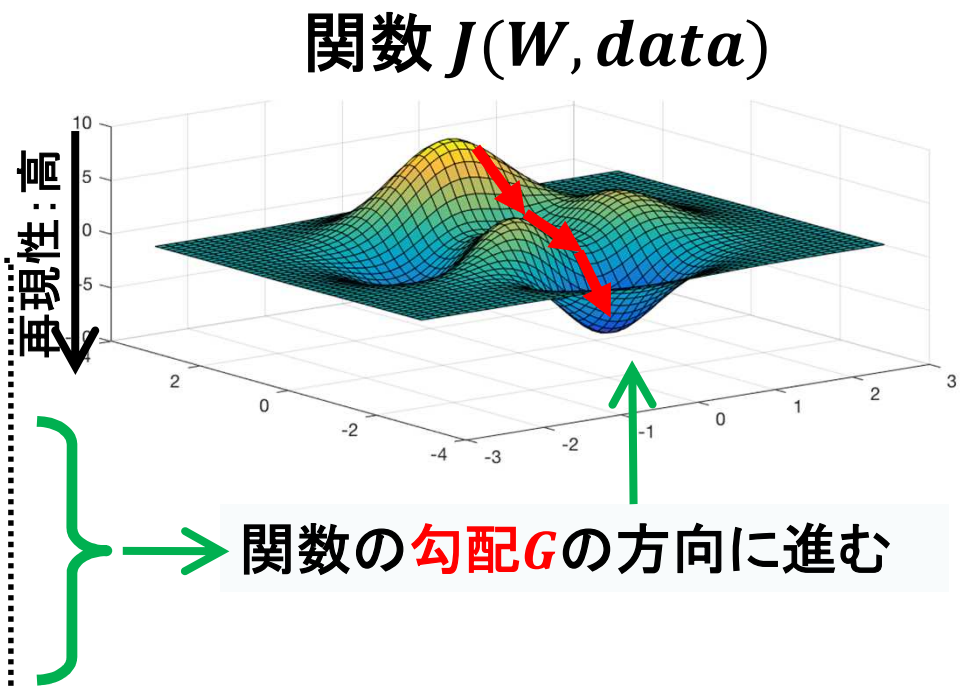


<u>ニューラルネットの学習</u>：　大量のデータ（＝入出力のペア）に対して、できる限りそれを再現するように重みベクトルWを調整

提案システムのため
準備2/2：確率的勾配降下法(SGD)

関数 $J(W, data)$：現在の重みWを用いたニューラルネットによるデータ再現性の指標

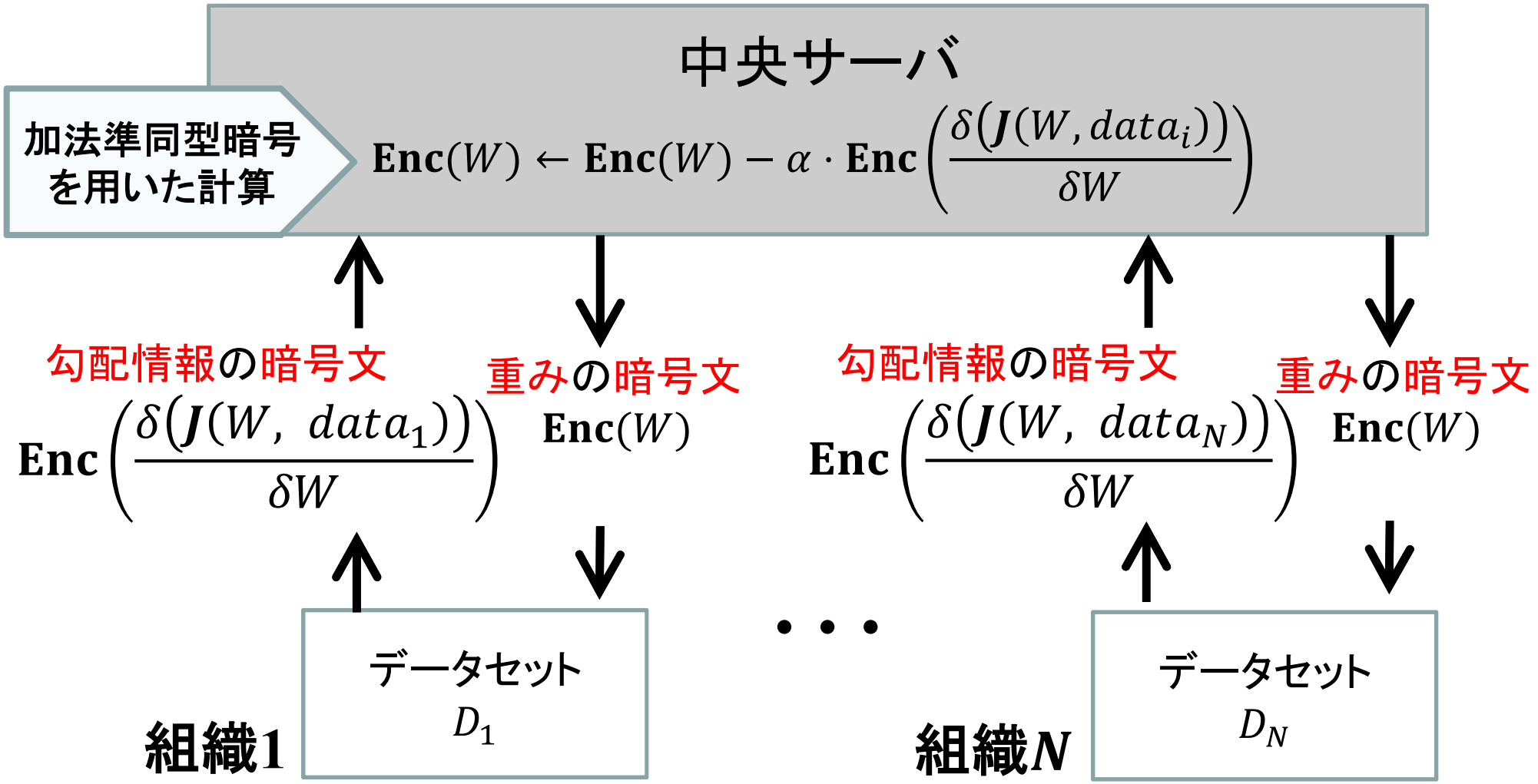
重みベクトルWをランダムに初期する
REPEAT
  FOR $data$ **IN Dataset:**
    $G \leftarrow \frac{\delta(J(W,\ data))}{\delta W}$　（勾配Gを計算）
    $W \leftarrow W - \alpha \cdot G$　（重みベクトルの更新）
  ENDFOR
UNTIL (good accuracy)



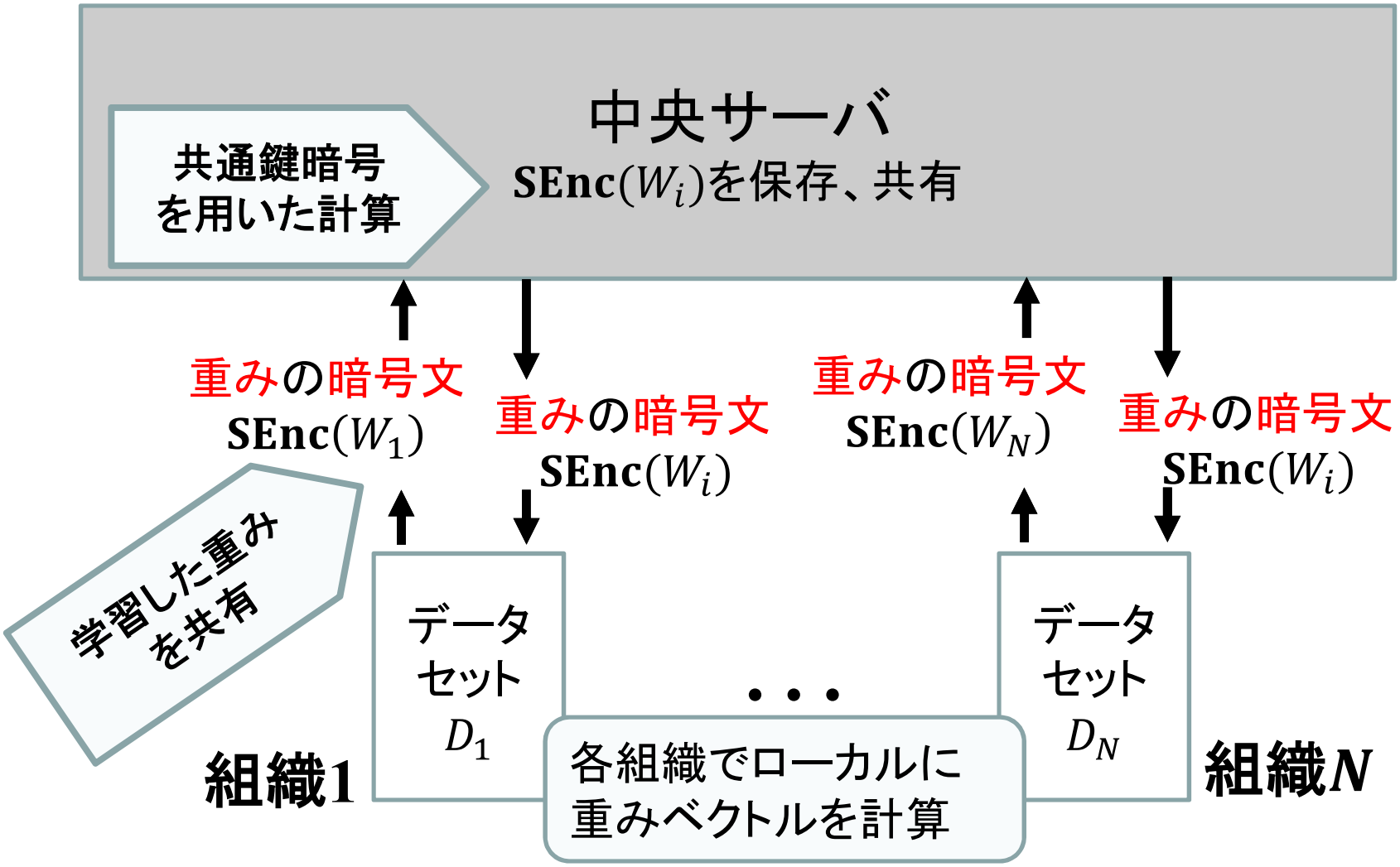
確率的勾配降下法(SGD)

NICT DeepProtect 1：加法型準同型暗号を用いたシステム

- 勾配情報を暗号化すれば、データ漏洩は防げる
- 中央サーバにおける計算は加算のみなので、暗号化したまま加算ができれば良い

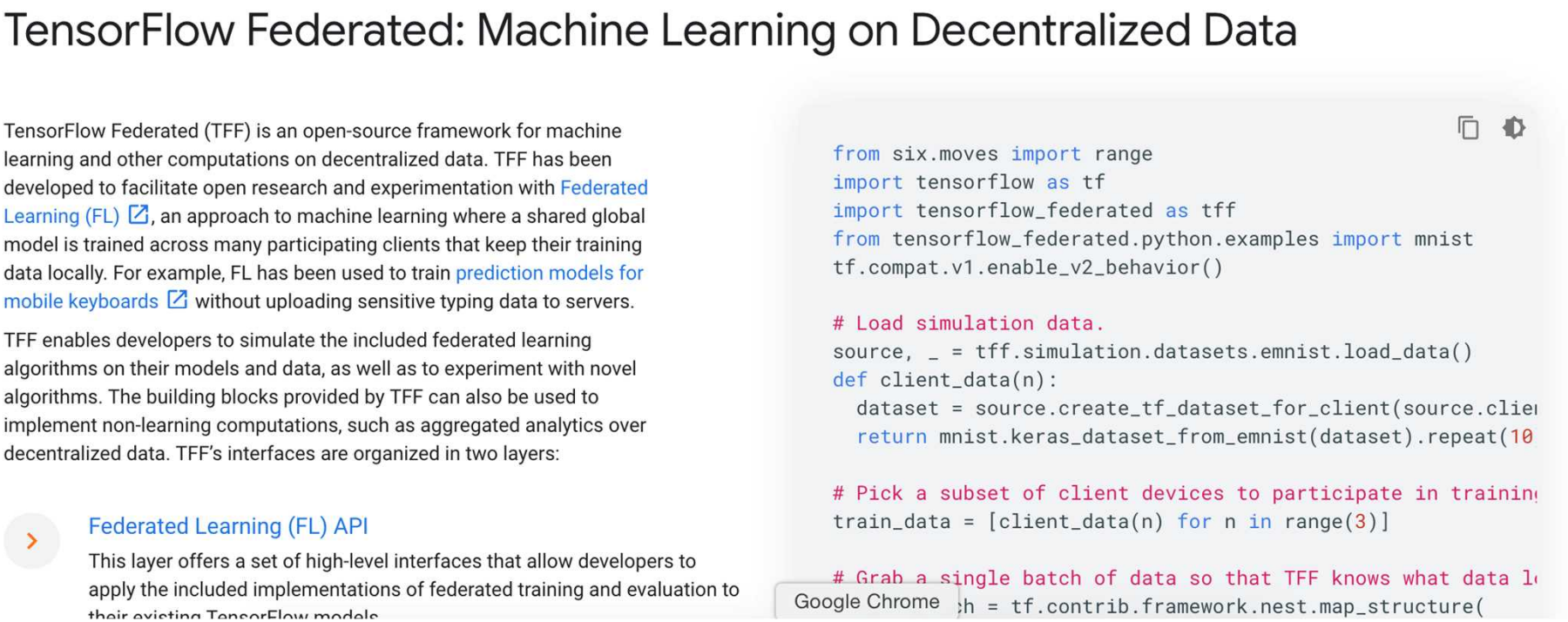


NICT DeepProtect 2：共通鍵暗号をシステム



類似技術：Google's Federated Learning（連合した学習）

- https://www.tensorflow.org/federated
- オープンソース: 2019年3月



TensorFlow Federated: Machine Learning on Decentralized Data

TensorFlow Federated (TFF) is an open-source framework for machine learning and other computations on decentralized data. TFF has been developed to facilitate open research and experimentation with Federated Learning (FL), an approach to machine learning where a shared global model is trained across many participating clients that keep their training data locally. For example, FL has been used to train prediction models for mobile keyboards without uploading sensitive typing data to servers.

TFF enables developers to simulate the included federated learning algorithms on their models and data, as well as to experiment with novel algorithms. The building blocks provided by TFF can also be used to implement non-learning computations, such as aggregated analytics over decentralized data. TFF's interfaces are organized in two layers:

Federated Learning (FL) API

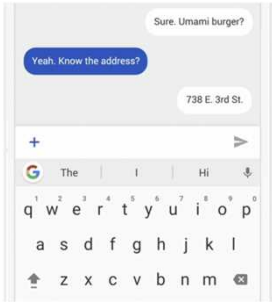
This layer offers a set of high-level interfaces that allow developers to apply the included implementations of federated training and evaluation to their existing TensorFlow models.

```
from six.moves import range
import tensorflow as tf
import tensorflow_federated as tff
from tensorflow_federated.python.examples import mnist
tf.compat.v1.enable_v2_behavior()

# Load simulation data.
source, _ = tff.simulation.datasets.emnist.load_data()
def client_data(n):
  dataset = source.create_tf_dataset_for_client(source.clie
  return mnist.keras_dataset_from_emnist(dataset).repeat(10

# Pick a subset of client devices to participate in trainin
train_data = [client_data(n) for n in range(3)]

# Grab a single batch of data so that TFF knows what data l
ch = tf.contrib.framework.nest.map_structure(
```

Google Chrome

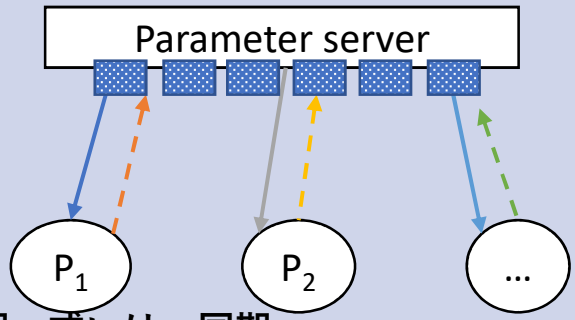
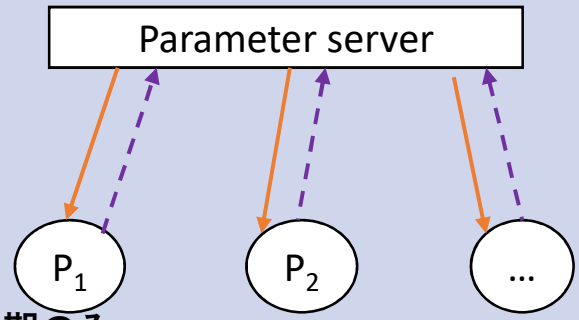
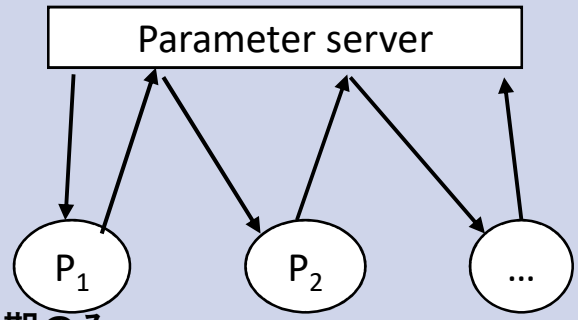
Google's Federated Learningの製品化

*"We're already using Federated Learning to improve several Google products. The Pixel first and second generation phones, for example, use Federated Learning to surface more accurate, useful **settings search** results so that people can find what that they're looking for faster. The Pixel has thousands of settings to adjust, from font size and brightness to app preferences and battery use. Different settings apply to different people and use cases, so personalizing users' experiences with machine learning can help people more easily find the one that they care about."*
https://ai.google/stories/ai-in-hardware/



Google Pixel 3
2018/11から
日本で
販売されている

通信の比較

	Shokri & Smartikov ACM CCS 2015 NICT DeepProtect 1	Googleの Federated Learning	NICT DeepProtect 2
通信の仕方	 **非同期、或いは、同期** ↑: 勾配の和 ↓:ニューラルネットワークの重み	 **同期のみ** ↑:ニューラルネットワークの重み ↓:ニューラルネットワークの重み	 **同期のみ** ↑:ニューラルネットワークの重み ↓:ニューラルネットワークの重み
応用	学習参加者P_iは、 （銀行や病院のような）組織	学習参加者P_iは、 スマートフォン	学習参加者P_iは、 （銀行や病院のような）組織

- R. Shokri, V. Shmatikov: *Privacy-Preserving Deep Learning*. ACM Conference on Computer and Communications Security 2015: 1310-1321
- B. McMahan et al.: *Communication-Efficient Learning of Deep Networks from Decentralized Data*. AISTATS 2017: 1273-1282
- **(DeepProtect1)** L. T. Phong et al. :*Privacy-Preserving Deep Learning via Additively Homomorphic Encryption*. IEEE Trans. Information Forensics and Security 13(5): 1333-1345 (2018)
- **(DeepProtect2)** L.T. Phong, T.T. Phuong: *Privacy-Preserving Deep Learning via Weight Transmission*, IEEE Trans. IFS, 2019, accepted

安全性と学習精度の比較

技術	中央サーバに対する安全性	中央サーバ＋不正の学習参加者に対する安全性	学習精度	暗号技術
Shokri and Shmatikov **(ACM CCS 2015)**	Onewayness under subset sum*	Onewayness under subset sum*	**非同期** SGD	TLS/SSL
NICTの DeepProtect 1	**Semantic security under encryption**	Onewayness under subset sum*	**非同期** SGD	TLS/SSL ＋ **準同型暗号**
Googleの 連合学習	Onewayness under subset sum	Onewayness under subset sum*	**同期** SGD	TLS/SSL
NICTの DeepProtect 2	**Semantic security under encryption**	Onewayness under subset sum*	**同期** SGD	TLS/SSL ＋ **共通鍵暗号**

*More generally, solving a system of non-linear equations where #unknowns > #equations

他データセット（医療や個人データ）を用いた実験（1/2）

UCI Dataset Name	Known accuracy (in [6], data privacy is preserved)	Known F-score (in [6], data privacy is preserved)	Percentage of label 0 (or label 1 if higher)	Our system accuracy (data privacy is preserved)	Our system F-score (data privacy is preserved)
Pima (diabetes)	80.70%	0.688525	64.29%	**85.06%**	**0.763636**
Breast Cancer	98.20%	0.962406	64.04%	**99.31%**	**0.989304**
Banknote Authentication	98.40%	0.984615	55.97%	**100.0%**	**1.0**
Adult Income	81.97%	0.526921	76.38%	**85.90%**	**0.664362**
Skin/NonSkin	93.89%	0.960130	79.53%	**99.95%**	**0.998655**

プライバシー保護
ロジスティックス回帰
(Aono et al., IEICE Trans. 2016)

プライバシー保護
ディープラーニング
(DeepProtect 2)

UCIデータセット
https://archive.ics.uci.edu/ml/index.php

他データセット（医療や個人データ）を用いた実験(2/2)

新技術説明会
New Technology Presentation Meetings!

プライバシー保護なし
のアルゴリズム

UCI Dataset Name	K	Our system accuracy (data privacy is preserved)	Our system F-score (data privacy is preserved)
Pima (diabetes)		85.06%	0.763636
Breast Cancer		99.31%	0.989304
Banknote Authentication		100.0%	1.0
Adult Income		85.90%	0.664362
Skin/NonSkin		99.95%	0.998655

	Algorithm	Error
1	C4.5	15.54
2	C4.5-auto	14.46
3	C4.5 rules	14.94
4	Voted ID3 (0.6)	15.64
5	Voted ID3 (0.8)	16.47
6	T2	16.84
7	1R	19.54
8	NBTree	14.10
9	CN2	16.00
10	HOODG	14.82
11	FSS Naive Bayes	14.05
12	IDTM (Decision table)	14.46
13	Naive-Bayes	16.12
14	Nearest-neighbor (1)	21.42
15	Nearest-neighbor (3)	20.35
16	OC1	
17	Pebls	

プライバシー保護
ディープラーニング
(DeepProtect 2)

(Aon

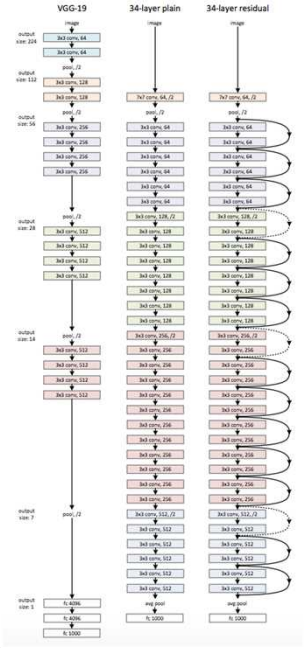
UCIデータセット
https://archive.ics.uci.edu/ml/i

プライバシー保護しながらも、
Adult Incomeデータセットにおいて、従来技術（プライバシー保護なし）の精度と比べても、ほぼ同等以上ある。

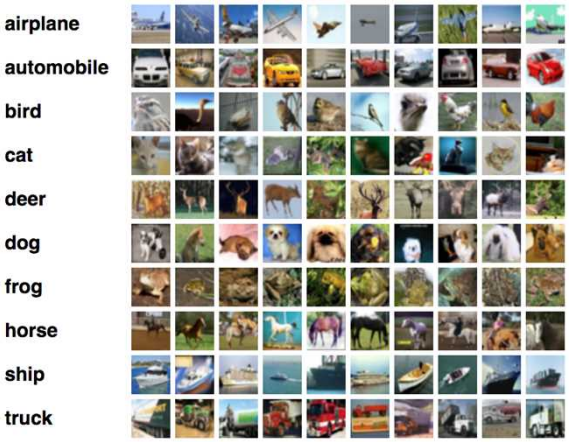
画像を用いた実験:
ResNet + CIFAR (10/100)

新技術説明会
New Technology Presentation Meetings!

ResNet



CIFARデータセット



https://arxiv.org/pdf/1512.03385.pdf
https://www.cs.toronto.edu/~kriz/cifar.html

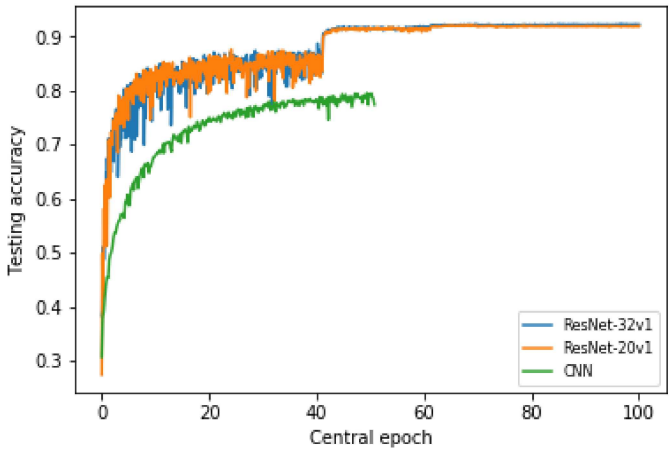
DeepProtect2:

暗号なしの学習時間は、暗号ありの学習時間とほぼ変わらない

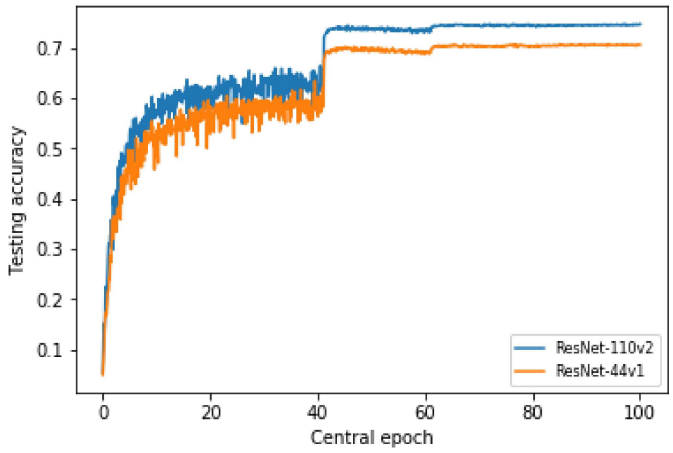
Network model, Dataset	$T^{(i)}_{original}$	$T^{(i)}_{uploadCtxt}$	$T^{(i)}_{downloadCtxt}$	$T^{(i)}_{enc}$	$T^{(i)}_{dec}$
CNN, CIFAR-10	37 (sec.)	0.056 (sec.)	0.056 (sec.)	0.130 (sec.)	0.06 (sec.)
ResNet-20v1, CIFAR-10	59 (sec.)	0.013 (sec.)	0.013 (sec.)	0.064 (sec.)	0.06 (sec.)
ResNet-32v1, CIFAR-10	87 (sec.)	0.023 (sec.)	0.023 (sec.)	0.066 (sec.)	0.06 (sec.)
ResNet-44v1, CIFAR-100	122 (sec.)	0.032 (sec.)	0.032 (sec.)	0.091 (sec.)	0.085 (sec.)
ResNet-110v2, CIFAR-100	322 (sec.)	0.165 (sec.)	0.165 (sec.)	0.179 (sec.)	0.138 (sec.)

In the table, the time for training over plain local data $T^{(i)}_{original}$ dominates the others.

(a) Experiments with CIFAR-10



(b) Experiments with CIFAR-100



想定される用途

- 金融分野において不正送金（振り込め詐欺等）の自動検知の精度向上に期待。
- 医療分野においても、プライバシーを保護したまま解析の精度を向上することが可能。
- プライバシー保護の目的外でも、データが分散するシナリオにおいて学習の通信量を削減する効果も可能。

実用化に向けた課題

現在、一つマシンでのシミュレーションやオープンデータの実験は開発済み。しかし、

a) 複数台のマシンの（効率かつ安定の）通信

b) 複数組織のコラボレーション（契約）

の２点は、未解決である。

企業への期待

- 未解決のa)については、GPUを持つマシンとマシンの通信技術により、克服できると考えている。
- 未解決のb)については、データを持つ企業や組織の協力により、克服できると考えている。
- 上記に関して、企業との共同開発を希望。
- また、深層学習を開発中の企業、組織横断の学習への展開を考えている企業には、本技術の導入が有効と思われる。

本技術に関する知的財産権（1）

- 発明の名称：サーバ、サービス方法
- 特許番号　　：6490429
- 出願人　　：国立研究開発法人
  情報通信研究機構
- 発明者　　：レ チュウ フォン、
  青野 良範、
  林 卓也、
  王 立華

本技術に関する知的財産権（2）

- 発明の名称：学習システム及び学習方法
- 出願番号　　：特願2018-001656
- 出願人　　：国立研究開発法人
  情報通信研究機構
- 発明者　　：レ　チュウ　フォン

お問い合わせ先

国立研究開発法人情報通信研究機構
イノベーション推進部門
技術移転コーディネータ　宇梶、橘田

ＴＥＬ　０４２－３２７　－　６９５０
ＦＡＸ　０４２－３２７　－　６６５９
e-mail　ippo@ml.nict.go.jp