情報系Winter Festa Episode 3

Genome Privacy CREST

Jun Sakuma, U. Tsukuba (privacy)
Koji Tsuda, U. Tokyo (stat.)
Ichiro Takeuchi, NITECH (ML)
Kunihiro Noboru, U. Tokyo (crypto)
Yoshiji Yamada, U. Mie (genetics)

目標: 機密性・プライバシ保護の障壁を取り除き、孤立したsmall dataをBIG DATAに

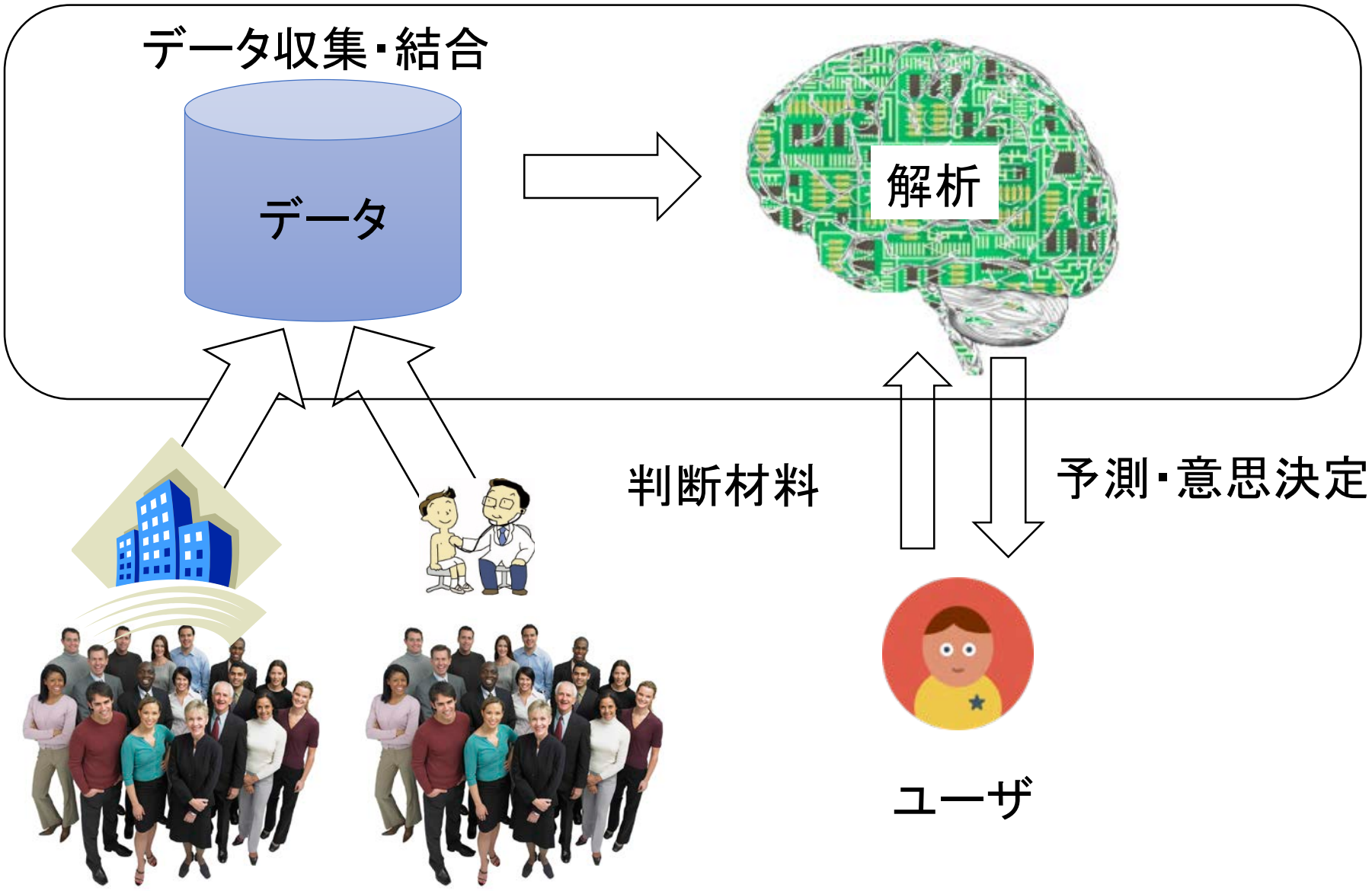
新サービス

新発見

孤立したSmall data

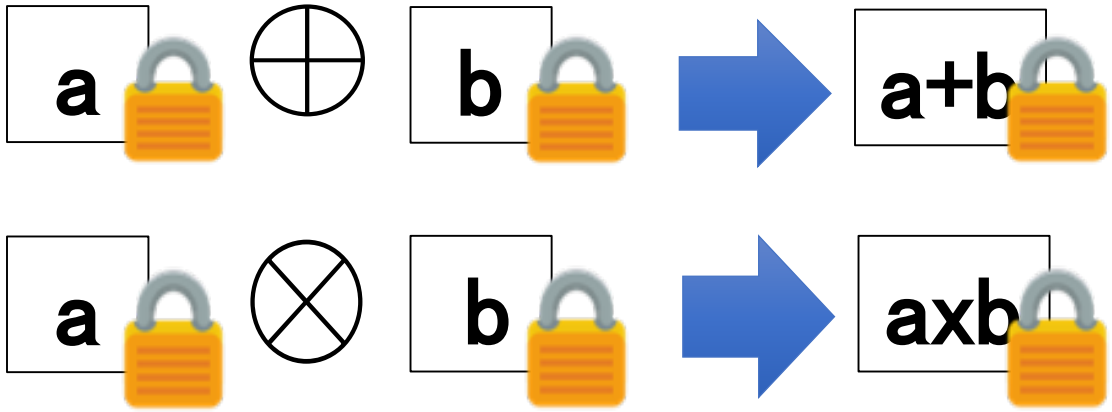
統計解析・機械学習におけるプライバシの問題

1. 個人情報は（同意なしには）結合できない



完全準同型暗号

- データを暗号化したまま加算・乗算
- 原理的には任意の演算を「データを見せずに」評価可能



Gentry Craig, “A fully homomorphic encryption scheme”, Doctoral dissertation, Stanford University, 2009

準同型暗号を用いた医療・ゲノムデータに基づく生活習慣病予防システム

遺伝要因によるリスク　　環境/臨床要因によるリスク

$$\mathrm{risk} = \boldsymbol{w}^T \boldsymbol{x} = w_1^G x_1^G + w_2^G x_2^G + \ldots + w_1^C x_1^C + w_2^C x_2^C + \ldots$$

SNP1が疾病に及ぼす影響

SNP1の有無

環境/臨床要因 (e.g. 年齢、性別、飲酒歴、喫煙歴、血液検査結果 etc)

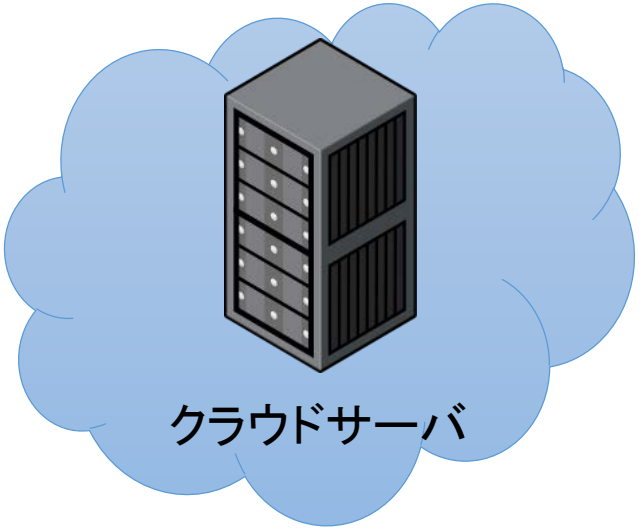
その臨床要因が疾病に及ぼす影響

暗号文上で評価可能

$$e(\boldsymbol{w}^T \boldsymbol{x}) = e(\sum_i w_i^G x_i^G + \sum_j w_j^C x_j^C)$$

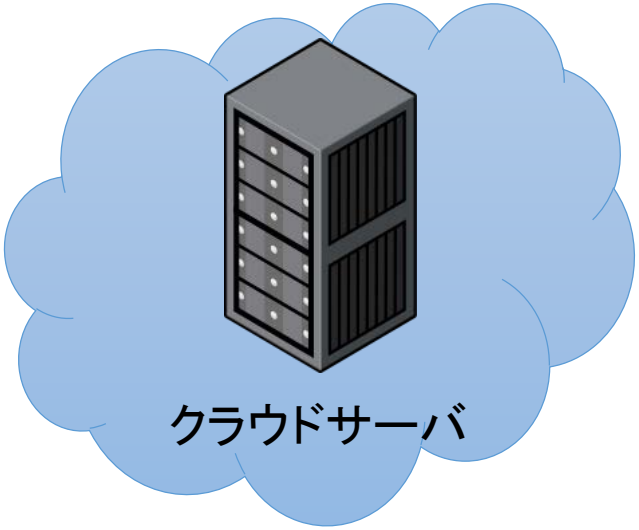
$$= \prod_i e(w_i^G x_i^G) \otimes \prod_j e(w_j^C x_j^C) \quad = \prod_i e(x_i^G)^{w_i^G} \otimes \prod_j e(x_j^C)^{w_j^C}$$

遺伝要因の暗号文　　臨床要因の暗号文



診療所

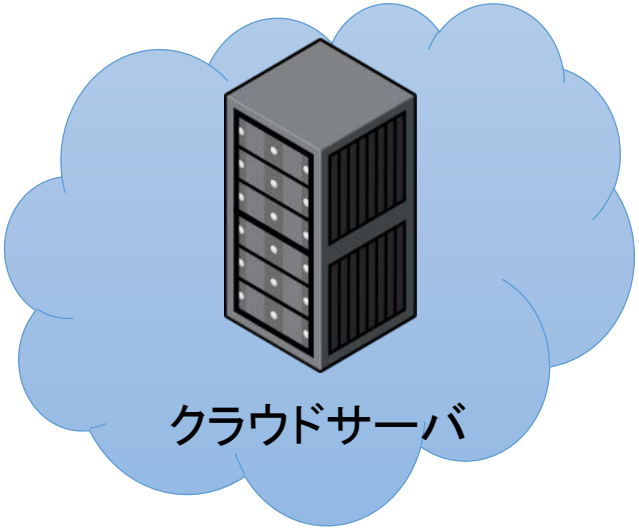
遺伝子検査機関



診療所

血液サンプル

遺伝子検査機関



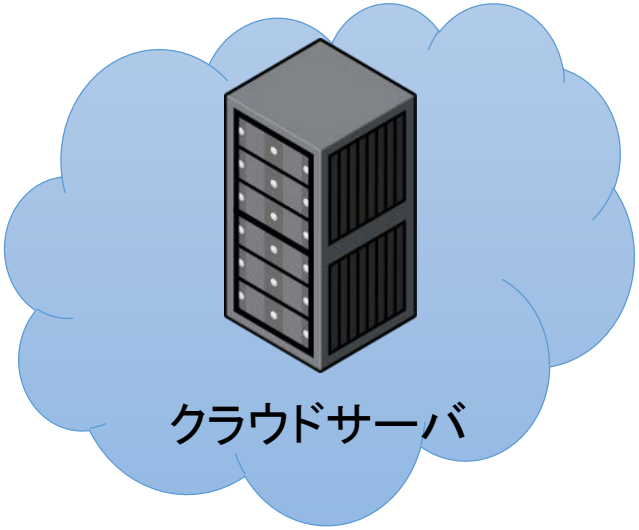
暗号化された
遺伝データ

遺伝子検査機関

診療所

SNPデータ

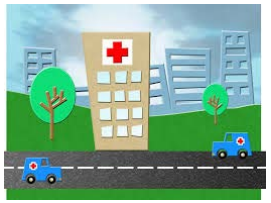
血液サンプル

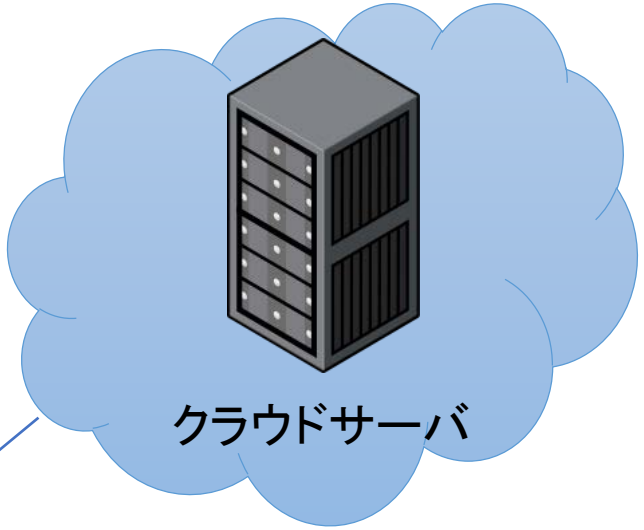


暗号化された臨床データ

診療所

遺伝子検査機関



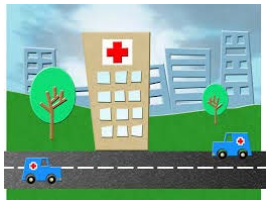


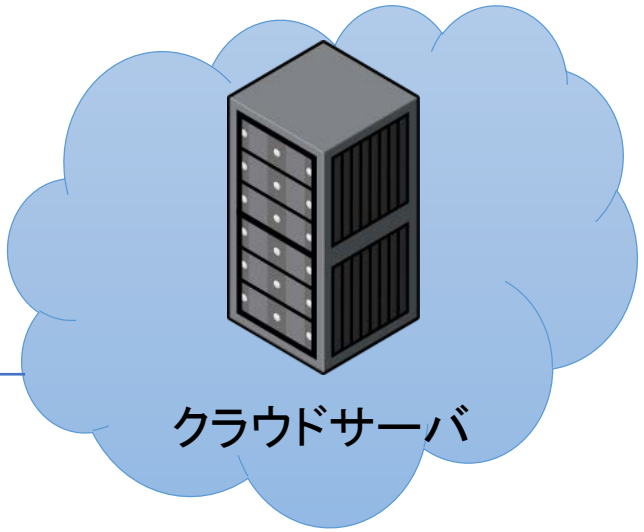
クラウドサーバ

診療所

暗号化された
分析結果のレポート

遺伝子検査機関





後日総合病院へ…

現在，2病院で
実証実験中

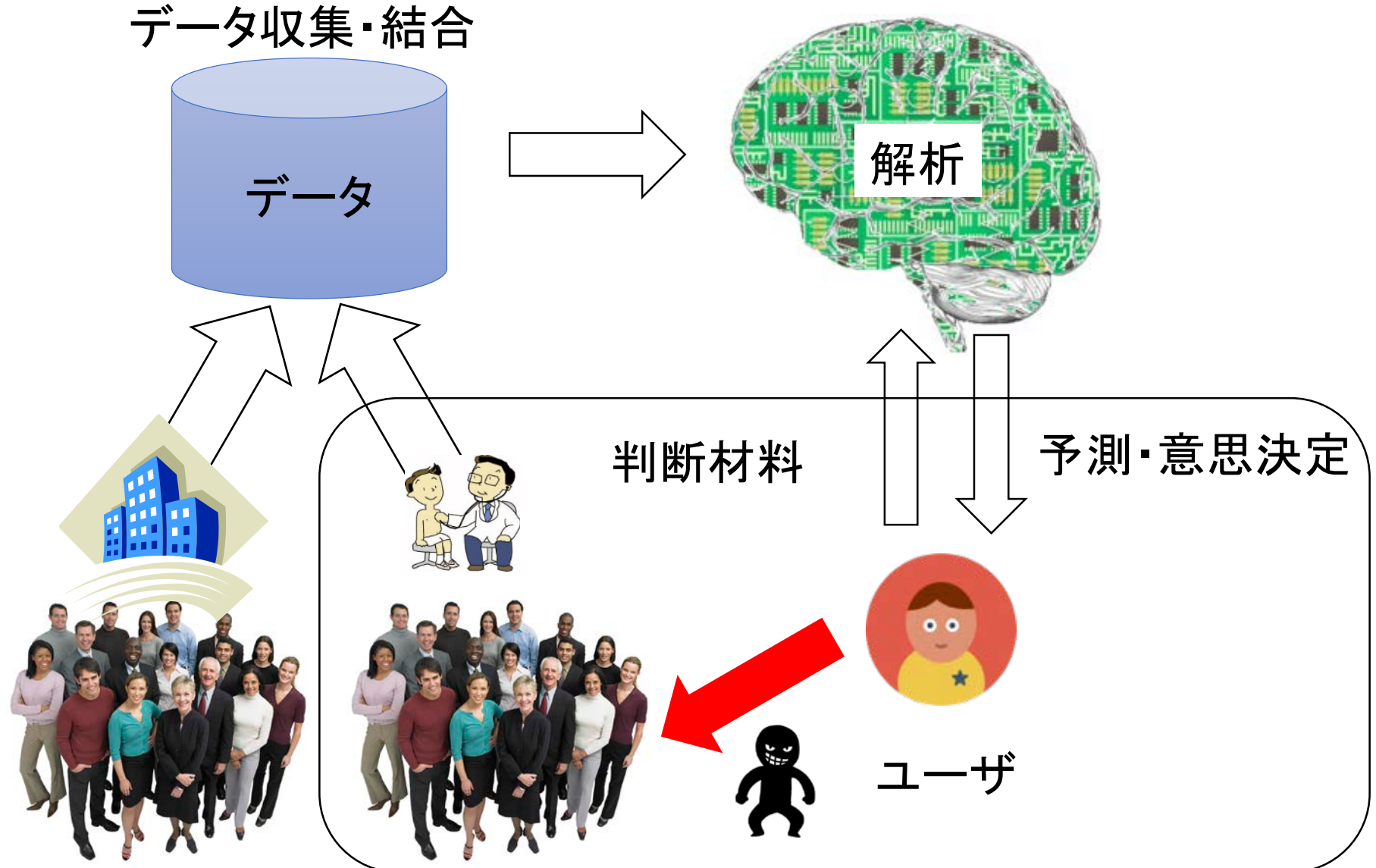
遺伝子検査機関

診療所

暗号化された
分析結果のレポート

統計解析・機械学習におけるプライバシの問題

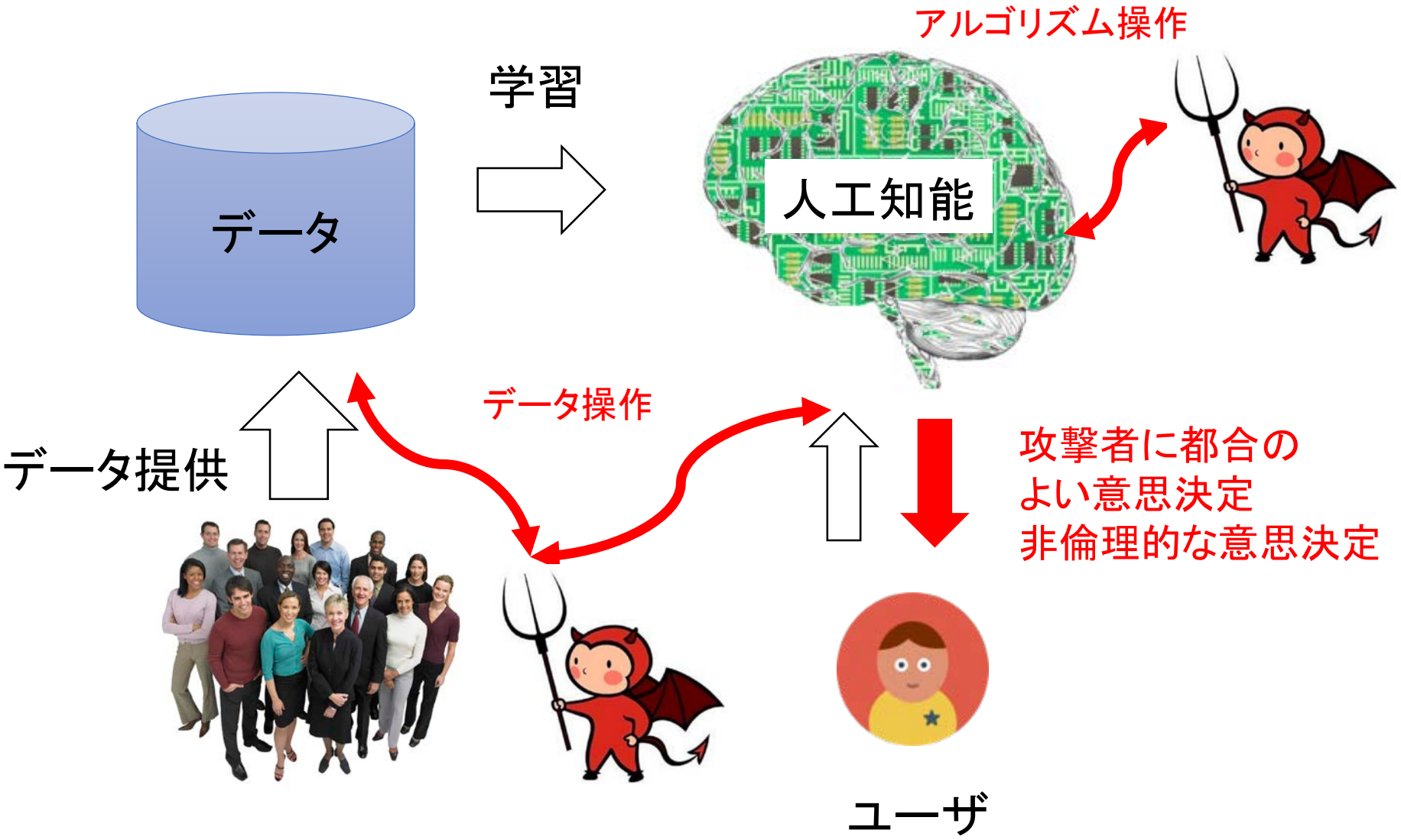
2. 公開統計情報から、個別入力が推測可能か評価できない



出力から入力が推定されないためには？

- ランダム化による保護
  - 出力をノイズで汚せば、入力はバレない
  - カイ二乗検定の差分プライバシー[Kakizaki+, ICML '17]
- 区間化による保護
  - 出力を区間化すれば、入力はバレない
  - 列挙によるリスク評価[Kusano+, ASIACCS' 17]
  - BDDによる近似リスク評価, プレビュー@竹内(東大)
  - 区間化データのERM[Hanada+, AAAI' 18], プレビュー@花田(名工大)

AIセキュリティ・プライバシー@理研AIP

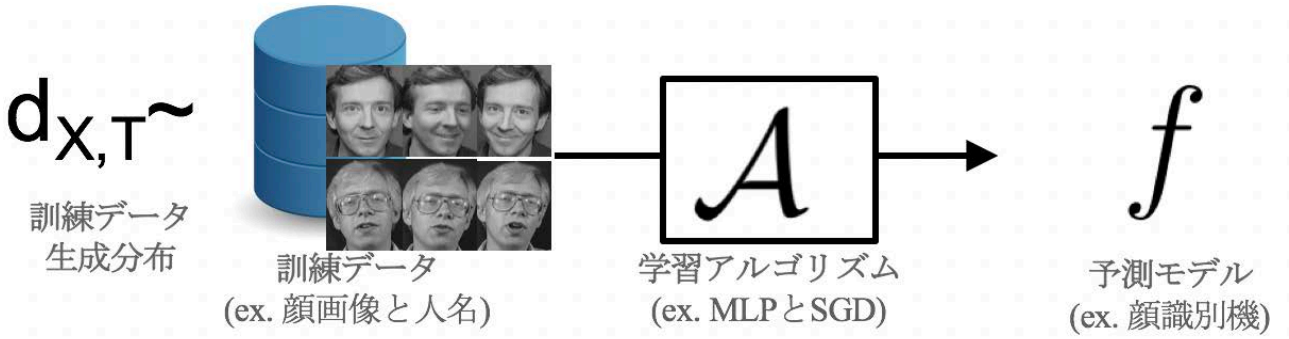


差別を起こさないバンディットの解析→プレビュー@福地(筑波)

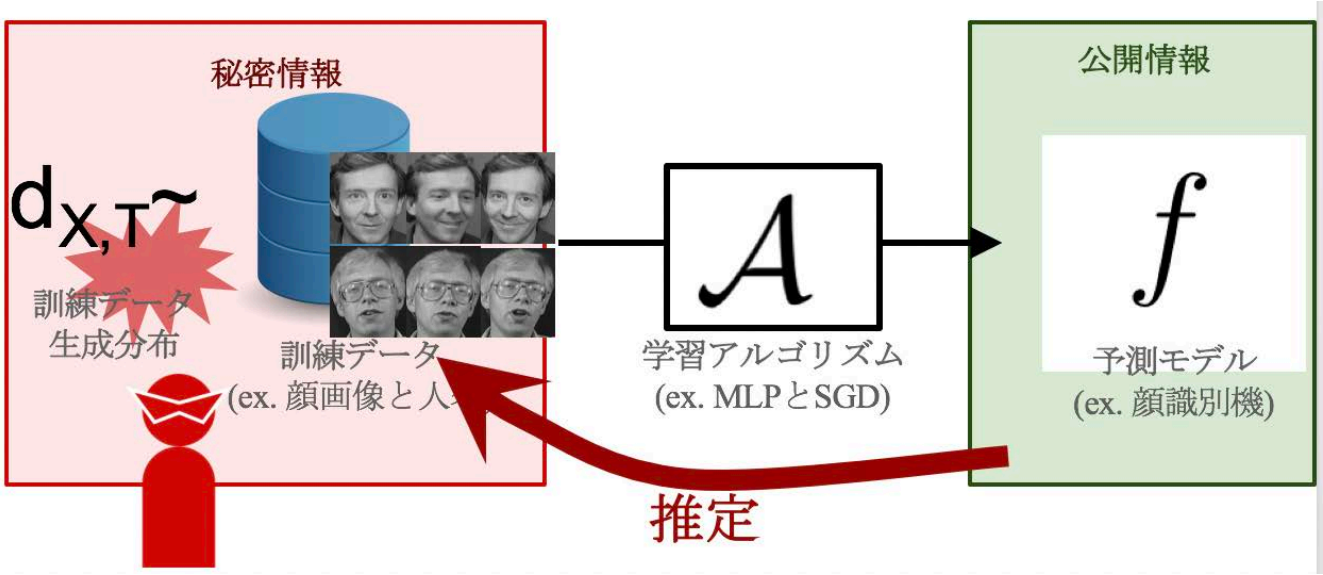
Calssifier-to-Generator攻撃

Kusano+, under review

- 通常の機械学習



- 分類器から訓練データを復元できるか？



畳み込みニューラルネットワークの秘密計算プロトコル

Setting:

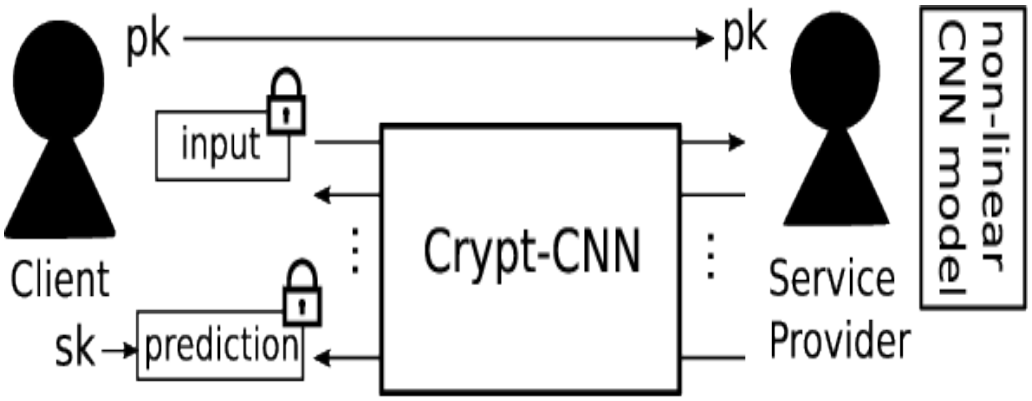
Both client's input and CNN model are private.

Objective:

To evaluate a CNN without disclosing any private information.

Keywords:

Homomorphic encryption, Garbled circuit, Secure computation



二分決定図を用いた，区間値公開時のプライバシ安全性の計算

CREST 佐久間プロジェクト　東京大学 竹内 聖悟

予測サービス

遺伝情報 → $f(x) \to y$ → 疾病罹患リスク

ユーザーA:
SNP** 1
SNP~~ 0
:

肥満 0.95
高血圧 0.21
:

攻撃者による入力値推定
$f^{-1}(y) \to x$

“ユーザーAはSNP**を持つ” プライバシリスク

安全な（推定されにくい）設計
（例: 数値ではなく区間値を使う）
安全性評価:
α −obscure安全[草野+2016]
目的: 高速な計算

α −obscure安全:

$$\alpha \geq |\Pr[X_i = k | Y = y] - \Pr[X_i = k]|$$

事前分布と事後分布の差から評価

事前分布 $\Pr[X_i = k]$

事後分布 $\Pr[X_i = k \mid Y = y] = \dfrac{\sum_{x \in \{x \mid x \in f^{-1}[\{y\}], x_i = k\}} \Pr[X=x]}{\sum_{x \in f^{-1}[\{y\}]} \Pr[X=x]}$

解の列挙

今回の対象:
SNP情報からの疾病罹患リスク予測
区間値 : $\mathrm{lb} \leq f(x) \leq ub$, 複数の疾患
⇨ $Ax \leq b$ の解 xの列挙
$x \in \{0,1\}^d, A \in Z^{m \times d}, b \in Z^m$
既存手法: 二分決定図(BDD)の利用
提案: より効率的な計算手法

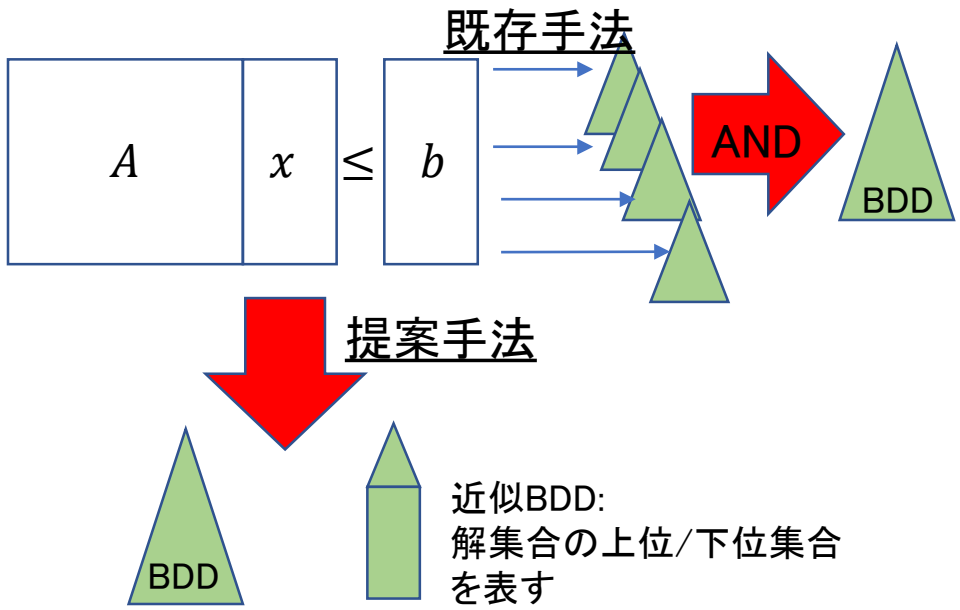
二分決定図を用いた，区間値公開時のプライバシ安全性の計算

	Behle	提案法	提案法（近似）
計算量			

BDDを用いた，解の列挙:

- 既存手法[Behle 2007]
  - 個々の不等式のBDDを構築，AND演算 ⇨ 解
  - AND演算は重い: 入力BDDサイズの乗算
- 提案手法1: **一度に全不等式を扱う**
  - 関連: フロンティア法 [Kawahara+2017]
- 提案手法+: **近似BDDの構築**
  - Relaxed, Restricted BDD [Bergman+2013]

W: BDD幅, w: 近似のパラメータ, BDD幅

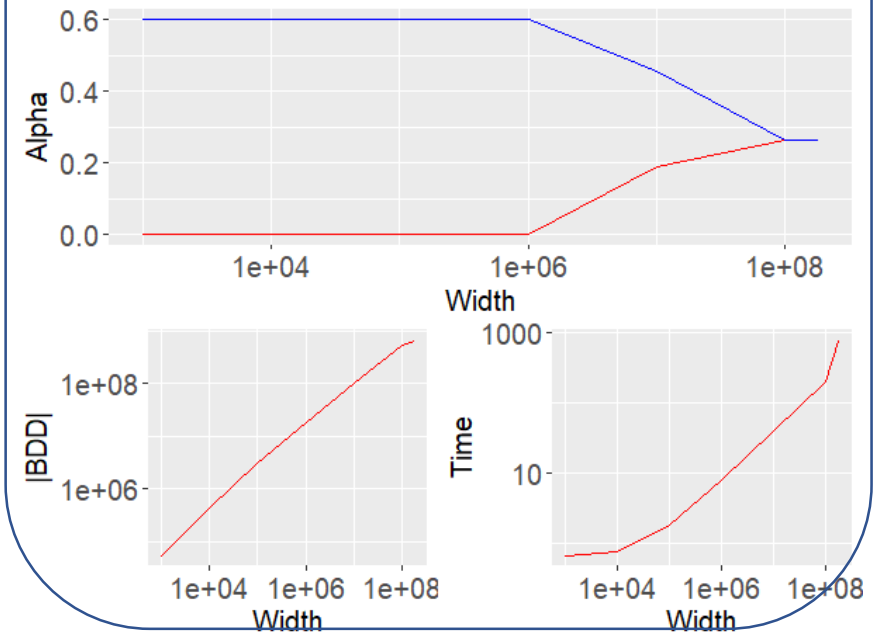


実験結果:

$m = [9,18], d = 99$

既存手法: ≥ 24時間 で解けず.

提案手法: ≤ 1,000秒



結果的公平な文脈付きバンディット学習

福地 一斗 (筑波大), 佐久間 淳 (筑波大/JST/理研)

文脈付きバンディット

例) 広告推薦

クリック率が上がるように
広告配信を最適化したい

公平性

例) (Sweeny 13)の差別的広告に関する報告

アフリカ系の名前

ネガティブな広告

ヨーロッパ系の名前

中立的な広告

センシティブ属性(人種など)に依存しない
広告配信をおこないたい

$$\max \text{累積報酬 sub to 不公平スコア} \leq \eta$$

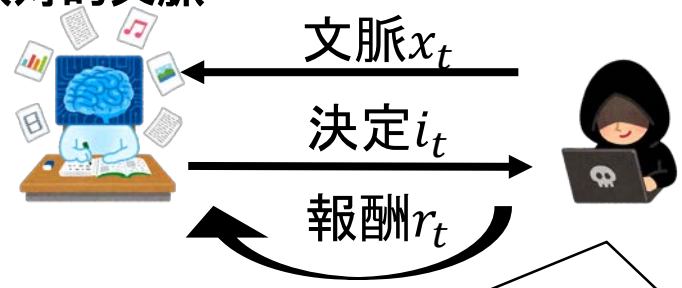
結果的公平な文脈付きバンディット学習

福地 一斗 (筑波大), 佐久間 淳 (筑波大/JST/理研)

研究目的

公平性を保証した文脈付きバンディットアルゴリズムの開発・解析

敵対的文脈

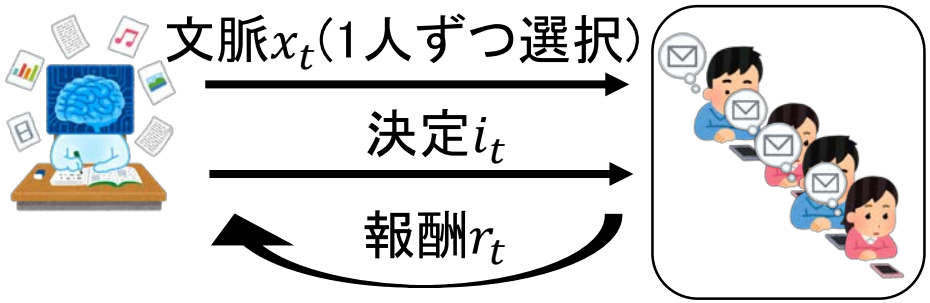


累積報酬が下がるようにx_tを決定

敵対的文脈は難しすぎて解けないことを示す

分配的文脈

登録ユーザ集合L



- **列線系レグレットを得られるアルゴリズムを提案・解析**
- **一定条件下における最適性を証明**