

PWS勉強会(2019/09/02)

匿名加工コンテストPWSCupの ルールや技術説明

村上 隆夫 (産総研)

PWSCup2019

▶ 特徴

- ▶ 位置情報コンテスト(我々の知る限り, 世界的に見ても本コンテストが初)
- ▶ ID識別とトレース推定の2軸での評価(両者の相関関係を明らかにする)
- ▶ 部分知識攻撃者モデル(提供先事業者が攻撃者と仮定)

	2015	2016	2017	2018	2019
データセット	疑似マイクロデータ (世帯消費額)	UCI Dataset "Online Retail" (購買履歴)			位置情報



【超重要】エントリーは下記HPから. まだ間に合います! (エントリー×:9/3)
<http://www.iwsec.org/pws/2019/entry/entry.html>

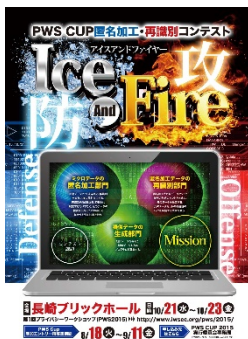
PWSCupの目的

▶ 目的1: 制度面での目的

- ▶ 改正個人情報保護法で「匿名加工情報」が定義されたものの、優れた匿名加工の方法が不明確. これを明確にする
 - ▶ 但し, 改正個人情報保護法における匿名加工情報の基準と, 本コンテストにおける匿名加工の安全性基準は異なる(例:前者は一般人基準, 後者は専門家レベル)
 - ▶ 両者の関係の明確化は今後の課題. ここでは便宜上, 「匿名加工」という言葉を使う
- ▶ 将来的な法制度の在り方を議論する上で, 参考となるような知見を提供する

▶ 目的2: 技術面での目的

- ▶ どのような匿名加工(or プライバシー保護)技術が良いのかを明らかにする



目次

PWSCup2019の概要

(位置情報コンテストの流れ, ID識別とトレース推定, 部分知識モデル)

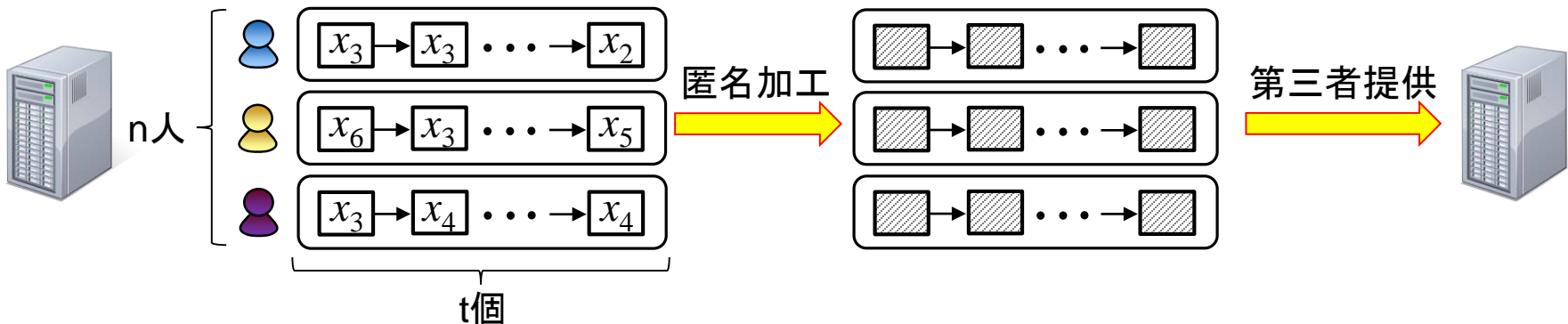
PWSCup2019の詳細

(データセット, 有用性指標, 安全性指標)

PWSCup2019: 位置情報コンテスト

概要

- ▶ LBS (Location-based Service) プロバイダーがトレース (移動履歴) を匿名加工して第三者提供する. そのときの有用性と安全性を競う
- ▶ 前提条件:
 - ▶ 元データは **nt個 (n人 x 時刻t個分) の位置**
 - ▶ nt個の各位置情報を加工する (ユーザが不明なダミートレースの追加は×)
- ▶ 匿名加工:
 - ▶ 位置情報の加工 (ノイズ, 一般化, 削除) + 仮名化 (トレースのシャッフル)
- ▶ 有用性指標:
 - ▶ ノイズ付与前後のユークリッド距離を基にした有用性
- ▶ 安全性指標:
 - ▶ ID識別: ID識別率を基にした安全性
 - ▶ トレース推定: 推定位置と実際の位置とのユークリッド距離を基にした安全性

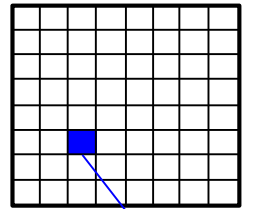


PWSCup2019: 位置情報コンテスト

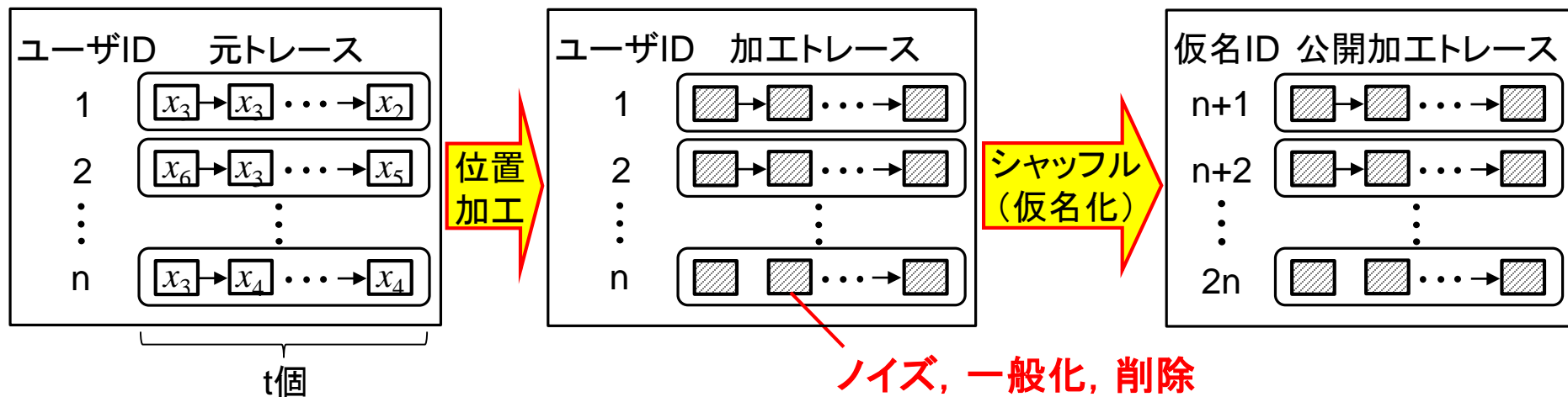
匿名加工 (Anonymization)

- ▶ n人のユーザに対して, 時刻t個分のトレースがあるとする
- ▶ 位置加工 (Obfuscation): 各位置情報を以下のように加工する
 - ▶ ノイズ付与 (例: $x_1 \Rightarrow x_2$)
 - ▶ 一般化 (元データ含む) (例: $x_1 \Rightarrow \{x_1, x_2, x_4\}$)
 - ▶ 一般化 (元データ含まない) (例: $x_1 \Rightarrow \{x_2, x_3, x_4\}$)
 - ▶ 削除 (例: $x_1 \Rightarrow \emptyset$)
- ▶ 仮名化 (Pseudonymization):
 - ▶ n個のトレースをランダムにシャッフルする

対象エリア

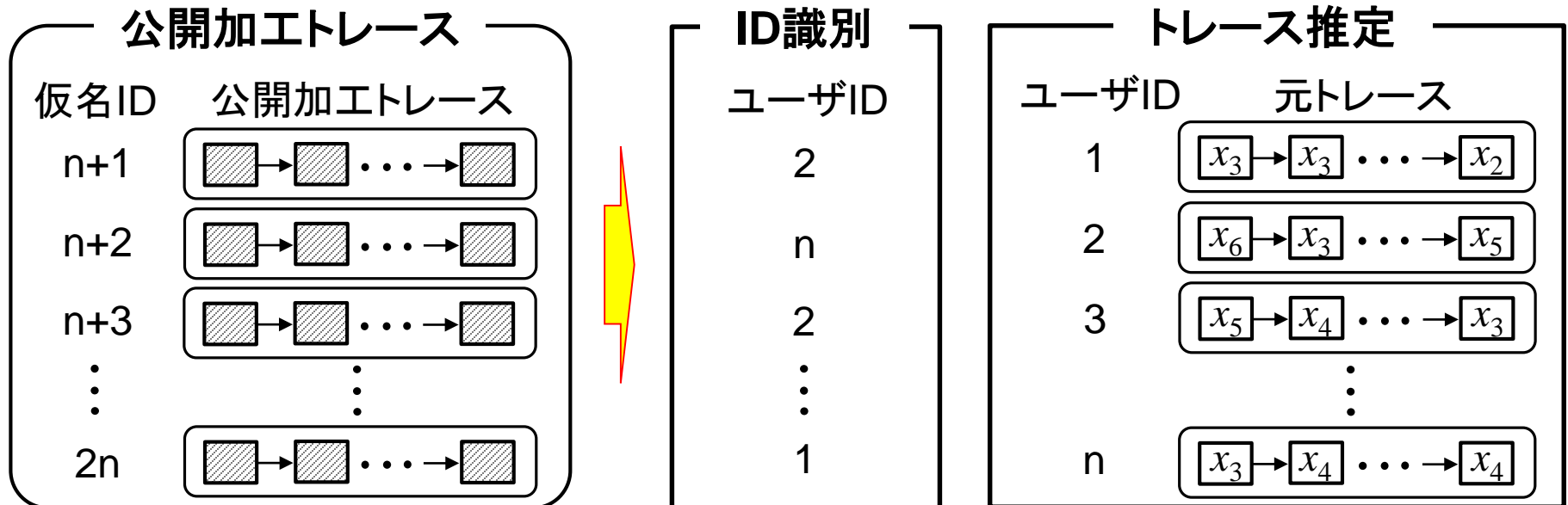


領域 x_i



PWSCup2019: 位置情報コンテスト

- ▶ ID識別 (ID Disclosure)
 - ▶ 各匿名加工エトレースに対して, n 人のユーザのうち誰かを当てる
 - ▶ 別名: 再識別
 - ▶ 出力: 1から n の自然数 \times n 行 (重複OK. 正解は1, 2, ..., n のpermutation)
- ▶ トレース推定 (Trace Inference)
 - ▶ nt 個 (n 人 \times 時刻 t 個分)の位置を推定する.
 - ▶ 別名: トラッキング攻撃 (元トレースを復元する攻撃) [Shokri+, S&P11]
 - ▶ 出力: nt 個の位置情報



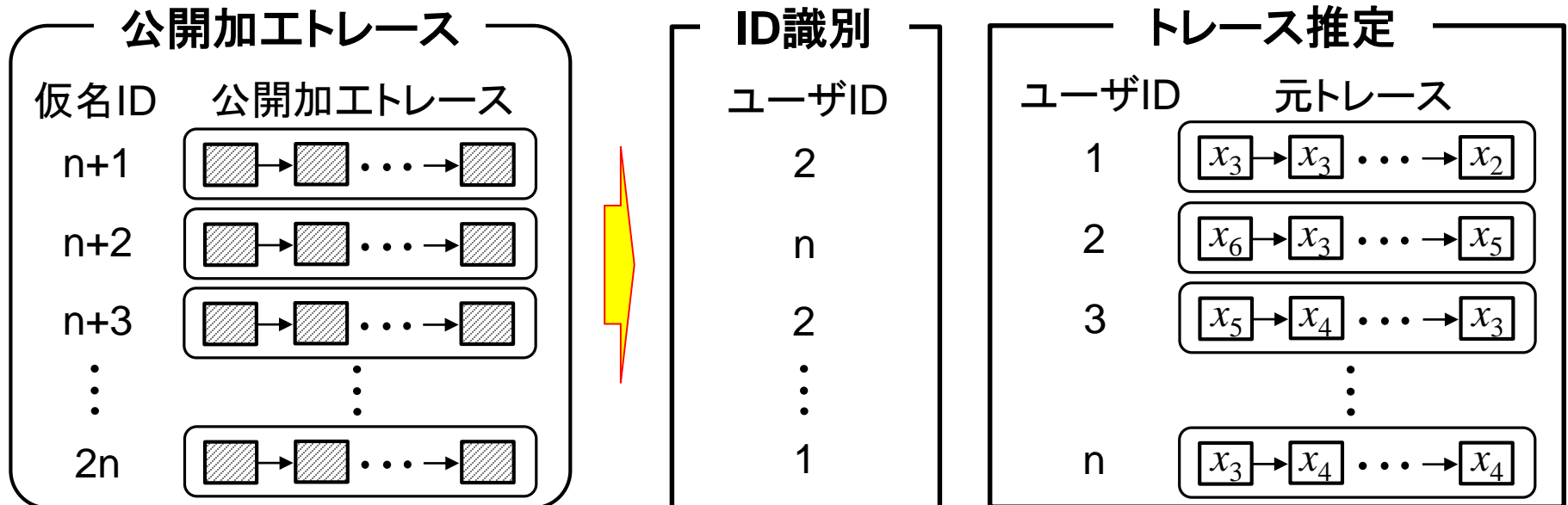
なぜID識別とトレース推定なのか？

▶ ID識別

- ▶ 仮名化トレースに対しては、「ユーザ \leftrightarrow 元の位置情報」の紐付けを行う
- ▶ 位置情報まで加工されたトレースに対しては、紐付けできるとは限らない

▶ トレース推定

- ▶ 位置情報まで加工された場合でも、「ユーザ \leftrightarrow 元の位置情報」の紐付けを行う



なぜID識別とトレース推定なのか？

- ▶ 現在の法律
 - ▶ ID識別のみをリスクとして考えており、トレース推定は対象外としている
- ▶ ID識別のみをリスクとした場合
 - ▶ K-匿名化が「任意の背景知識を持つ攻撃者」に対して安全（識別率 $\leq 1/K$ ）
- ▶ ID識別のみをリスクとして考える、というので本当に良いのか？
 - ▶ K-匿名化は、攻撃者にID識別を行うことなく**属性推定される**リスクが残る
 - ▶ 例：L-多様性論文[Machanavajjhala+, ICDE06]のhomogeneity attack
 - ▶ 同様に、ID識別に強いが、**トレース推定に弱い**加工例も存在する
 - ▶ サンプルプログラムを用いた評価実験で実証(⇒ PWSCup2019 HP)



PWSCup 2019(将来の法制度に向けて)

ID識別とトレース推定の2軸での評価(両者に対する安全性の関係を明らかに！)

ID識別とトレース推定の2軸での評価

▶ 概要

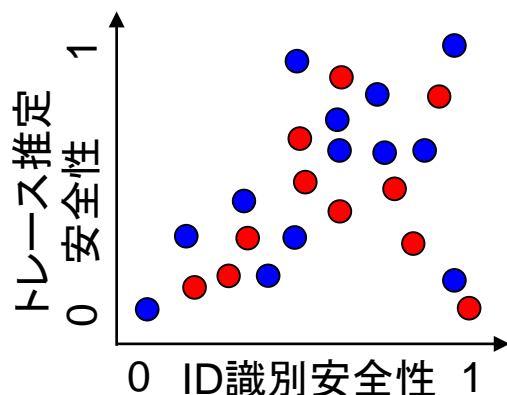
- ▶ 各チームに「ID識別対策用」、「トレース推定対策用」の2つの元トレースを配る
 - ▶ 加工の意図が明確になるよう、1つ目をID識別対策用、2つ目をトレース推定対策用とする

▶ 匿名加工フェーズ:

- ▶ 各チームは、最大2個の加工トレースを提出する(0個でも失格にはならない)

▶ ID識別・トレース推定フェーズ:

- ▶ 各チームは、他チームの公開加工トレースをID識別 and/or トレース推定する



- ID識別対策用
- トレース推定対策用

スコア: 0(悪い) - 1(良い)

〔有用性に関しては、**要求値**を設け、それを下回った加工データは無効とする〕

賞

▶ 総合優勝・総合2位・総合3位

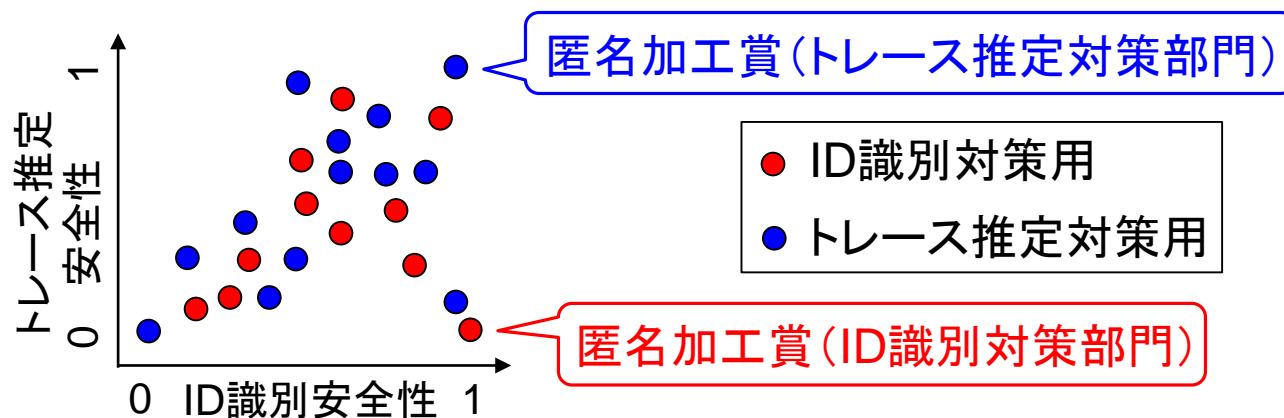
- ▶ 「ID識別対策用データのID識別安全性 + トレース推定対策用データのトレース推定安全性」の総和が最も大きい上位3チーム

▶ 匿名加工賞 (ID識別対策部門)

- ▶ ID識別対策用データのID識別安全性が最も大きい1チーム

▶ 匿名加工賞 (トレース推定対策部門)

- ▶ トレース推定対策用データのトレース推定安全性が最も大きい1チーム



賞

▶ リスク評価賞 (ID識別部門)

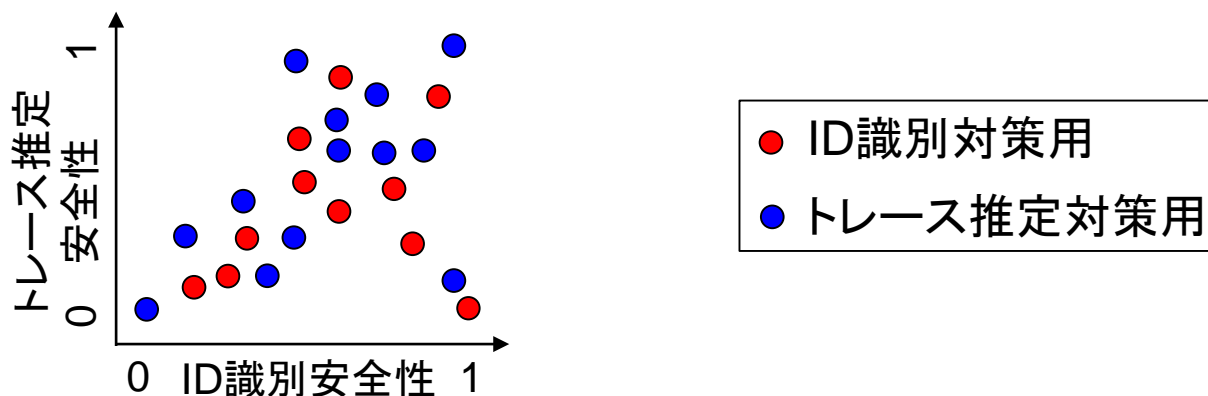
- ▶ ID識別安全性を最も下げた1チーム. **ID識別対策用**と**トレース推定対策用**の両データに対してID識別安全性を評価する(詳細はルール資料)

▶ リスク評価賞 (トレース推定部門)

- ▶ トレース推定安全性を最も下げた1チーム. **ID識別対策用**と**トレース推定対策用**の両データに対してトレース推定安全性を評価する(詳細はルール資料)



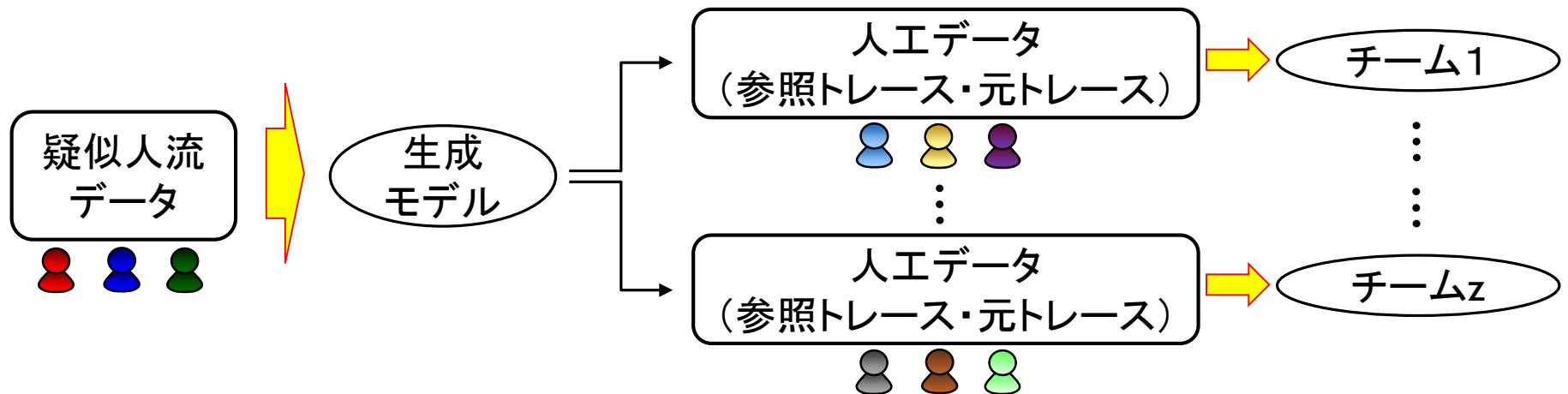
各データに対して, ID識別とトレース推定の2軸での評価が可能になる



(その他, プレゼンテーション賞あり)

最大知識モデル or 部分知識モデル？

- ▶ 最大知識モデル[Domingo-Ferrer+, PST15]
 - ▶ 攻撃者が元データを知っているというモデル(攻撃者 = 提供元事業者)
 - ▶ 【課題1】現実から乖離(元データを知っているのでID識別・トレース推定は不要)
 - ▶ 【課題2】位置情報は特異性が高く[Montjoye+, SR13], すぐID識別される
- ▶ 部分知識モデル(PWSCup2019)
 - ▶ 攻撃者が元データを知らないというモデル(攻撃者 = 提供先事業者)
 - ▶ 各チームに異なる(仮想的な)ユーザの人工データを配布する
 - ▶ 攻撃者は元トレースは知らないが, 参照トレースを知っていると仮定

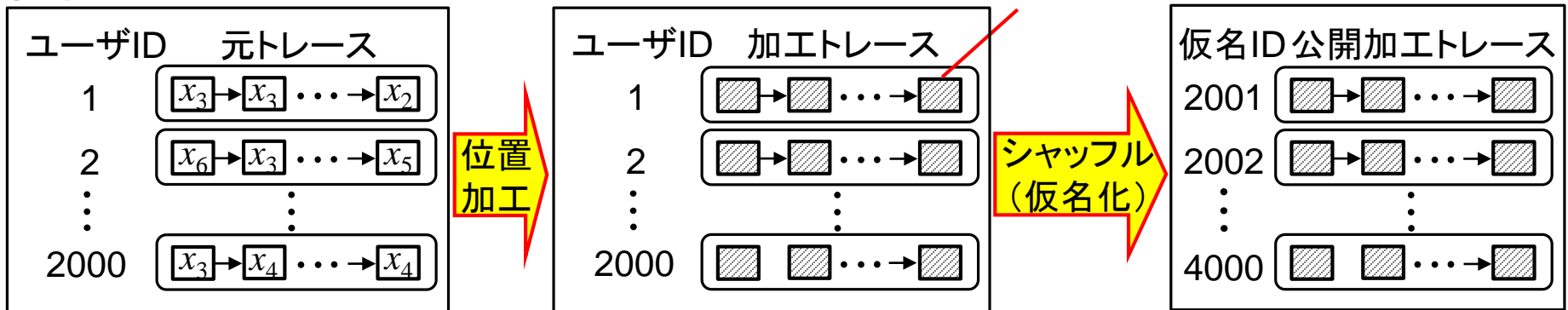


部分知識モデルのコンテスト

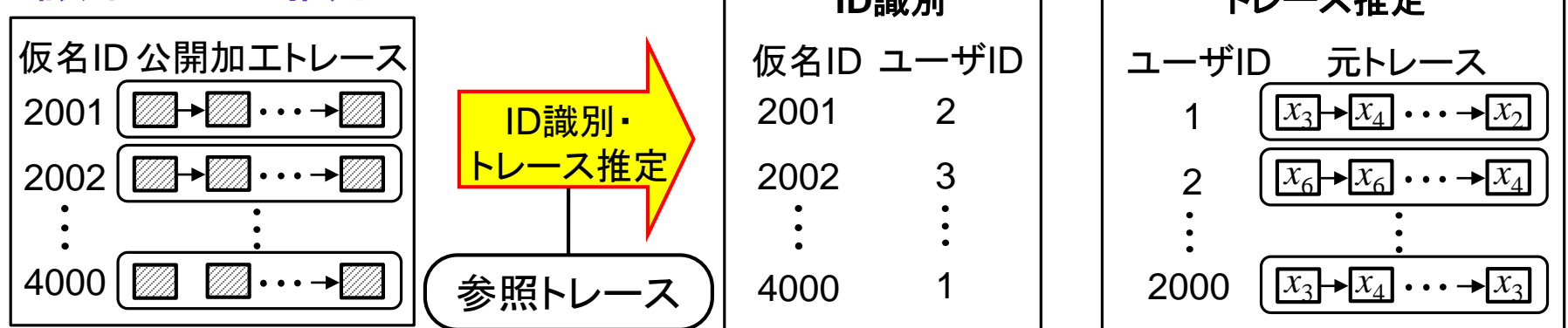
全体図

- 各チームは元トレースの位置情報を加工する(シャッフルは運営側で行う)
- 各チームは他チームに対して、参照トレースを基にID識別・トレース推定する

匿名加工フェーズ



ID識別・トレース推定フェーズ



目次

PWSCup2019の概要

(位置情報コンテストの流れ, ID識別とトレース推定, 部分知識モデル)

PWSCup2019の詳細

(データセット, 有用性指標, 安全性指標)




データセット

▶ PWSCup2019用人工データ




- ▶ 疑似人流データ(オープンな人工データ)を基に, 生成モデルを学習する
- ▶ チーム毎・データセット毎に異なる仮想ユーザのトレースを生成する

i 番目のチーム

ID識別対策用データセット

仮想ユーザ	参照トレース	元トレース
	$x_1 \rightarrow x_2 \rightarrow x_1 \rightarrow x_1$	$x_1 \rightarrow x_3 \rightarrow x_2 \rightarrow x_1$
	$x_4 \rightarrow x_5 \rightarrow x_5 \rightarrow x_5$	$x_4 \rightarrow x_4 \rightarrow x_5 \rightarrow x_5$
	$x_3 \rightarrow x_3 \rightarrow x_2 \rightarrow x_4$	$x_3 \rightarrow x_4 \rightarrow x_4 \rightarrow x_4$

トレース推定対策用データセット

仮想ユーザ	参照トレース	元トレース
	$x_5 \rightarrow x_5 \rightarrow x_4 \rightarrow x_3$	$x_5 \rightarrow x_4 \rightarrow x_3 \rightarrow x_3$
	$x_2 \rightarrow x_2 \rightarrow x_4 \rightarrow x_3$	$x_2 \rightarrow x_3 \rightarrow x_2 \rightarrow x_4$
	$x_1 \rightarrow x_4 \rightarrow x_4 \rightarrow x_1$	$x_1 \rightarrow x_1 \rightarrow x_4 \rightarrow x_4$

疑似人流
データ



ランダムな
生成モデル



疑似人流データ

▶ 疑似人流データ(オープンな人工データ)

▶ エリア: 東京近郊(首都圏)

▶ 対象時期: 6日間(2013年の7/1, 7/7, 10/7, 10/13, 12/16, 12/22)

2014.07.31 データ活用事例

東京大学CSISとの研究活動成果としてSNS解析データを元とした「疑似人流データ」を無料公開

ナイトレイでは、東京大学空間情報科学研究センター(CSIS)の柴崎・関本研究室、マイクロジオデータ研究会と共同で進めていた研究活動の成果として、「疑似人流データ」を本日から無料で公開致します。

「疑似人流データ」では、当社が保有するSNSベースの地域解析結果(地域ごと人気施設、生活者の行動傾向等)を参考に、正確な道路ネットワークデータによる移動経路の補完や統計処理・独自の推定処理、ランダム化処理を行うことで、東京近郊(首都圏)の大まかな人の流れをオープンなCSVデータとして公開するという取り組みです。



<https://nightley.jp/archives/1954/>

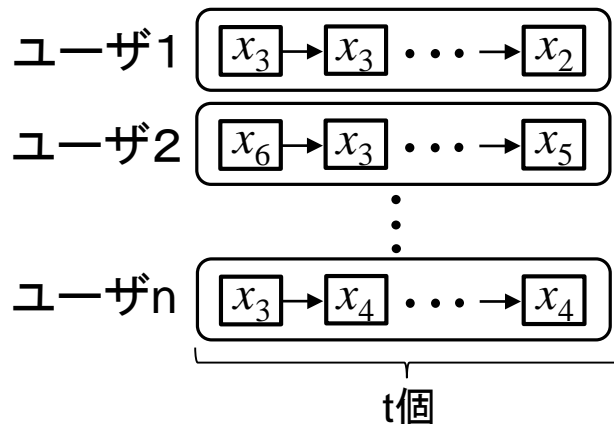
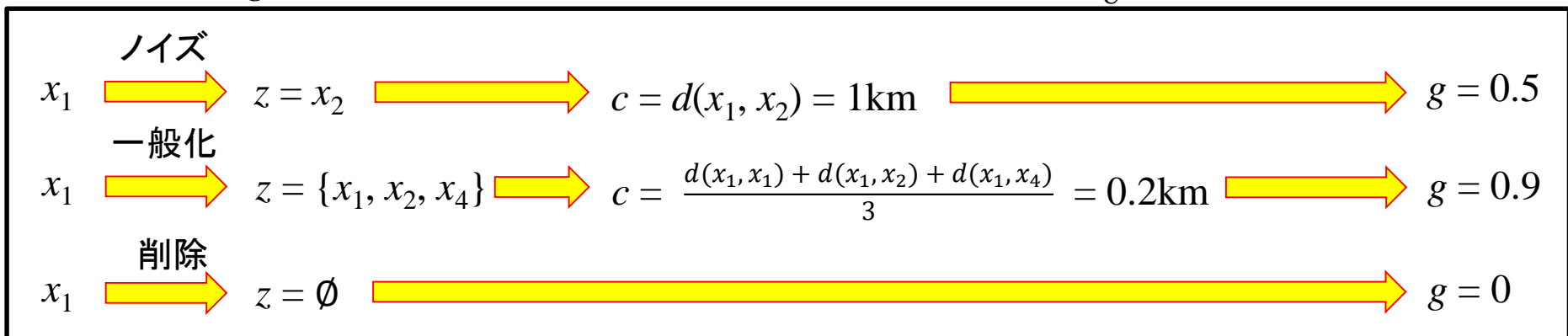
データセット

- ▶ PWSCup2019用人工データ(詳細)
 - ▶ ユーザ数: $n = 2000$
 - ▶ 位置情報数: $m = 1024$ (東京中心部を 32×32 の領域に分割)
 - ▶ トレースの長さ: 予備戦では $t = 40$ (8:00~17:59の2日分, 30分おき). 1, 2日目が参照トレース, 3, 4日目が元トレース. 本戦では日数変更の可能性あり
- ▶ 生成モデル(詳細はPWSCup HP)
 - ▶ マルコフモデルに基づく生成モデル(**詳細は非公開**). 以下の特徴を持つ
 - ▶ 人口分布の保存: 1時間毎の人口分布が疑似人流データに近い
 - ▶ 遷移行列の保存: 1024×1024 の遷移行列が疑似人流データに近い
 - ▶ 家のモデル化: 各ユーザは朝に高い確率で自身の家の領域にいる

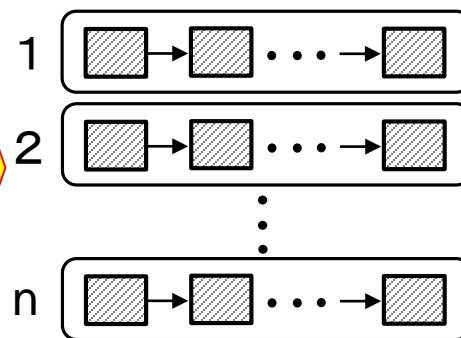
有用性指標

▶ 有用性

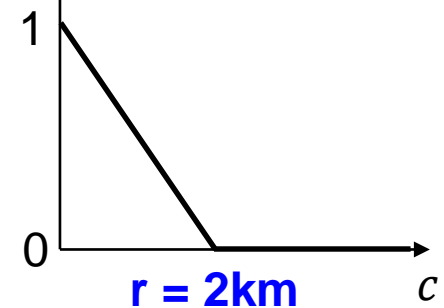
- ▶ 加工するほど下がり、一定以上加工すると完全に失われる(汎用性を考慮)
- ▶ nt 個の位置情報のそれぞれに対して、以下のスコア g を計算
 - ▶ Step 1. ノイズ付与前後の位置 x, z のユークリッド距離 $d(x, z)$ の平均 c を計算する
 - ▶ Step 2. c をスコア g (0:悪い, 1:良い) に変換する(削除に対しては $g = 0$)
- ▶ スコア g の nt 個の位置情報に対する平均を、有用性 s_U とする



位置加工



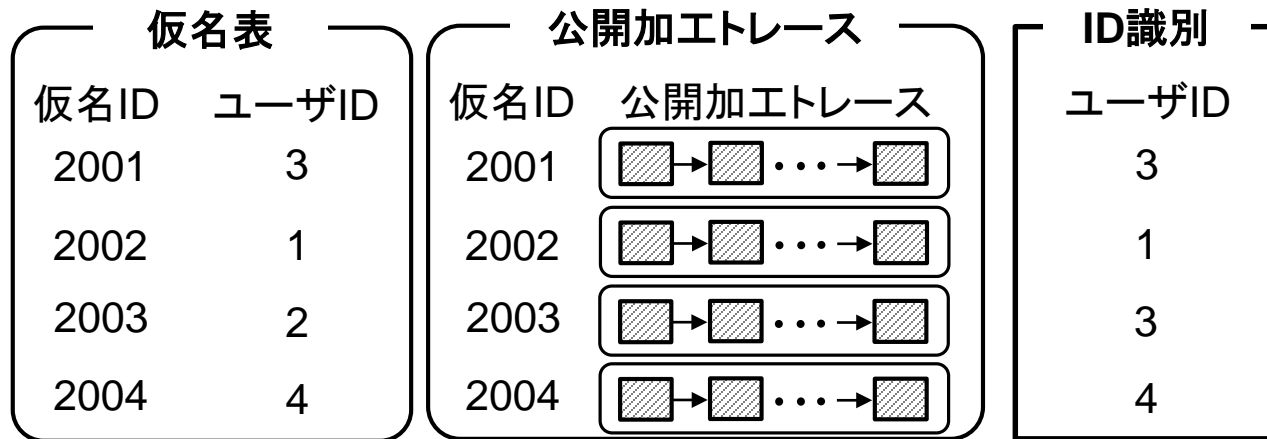
スコア g



安全性指標 (ID識別)

▶ ID識別安全性

- ▶ ID識別安全性 $s_I = 1 - \text{ID識別率}$ (0:悪い, 1:良い)



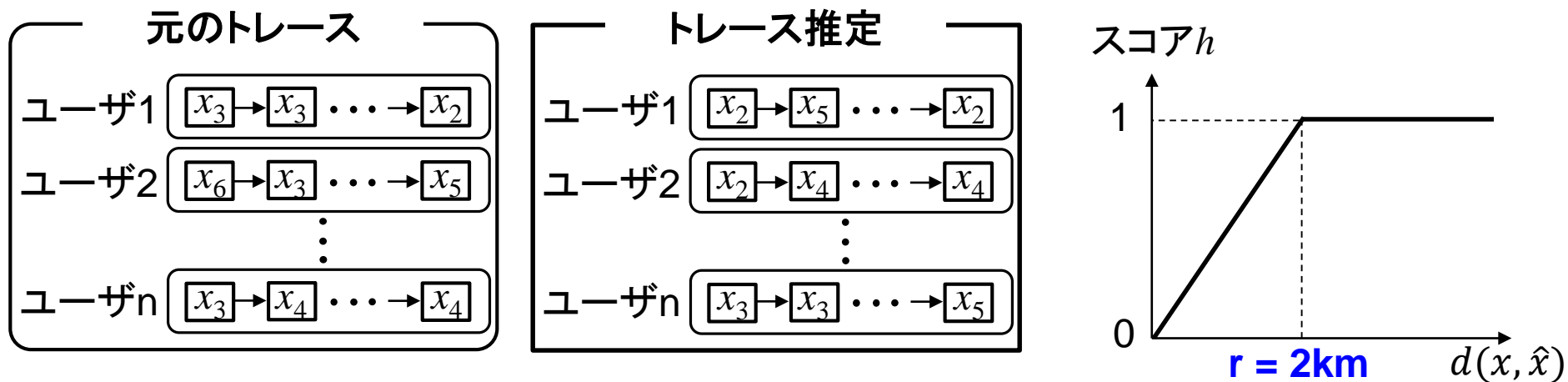
ID識別率 = $3/4 = 0.75$

ID識別安全性 $s_I = 0.25$

安全性指標(トレース推定)

▶ トレース推定安全性

- ▶ 実際の位置 x と推定位置 \hat{x} とのユークリッド距離 $d(x, \hat{x})$ をスコア h に変換
- ▶ h の nt 個の位置情報に対する重み付け平均を, トレース推定安全性 s_T とする



▶ 重み付け平均

- ▶ 疑似人流データから, 病院カテゴリーのPOIを含む領域(計37個)を抽出
- ▶ 病院領域(通称:ドラ)に対しては重みを10倍にして平均をとる

まとめ

▶ PWSCup2019

- ▶ 部分知識モデルでの位置情報コンテスト. ID識別とトレース推定の2軸での評価を行う(両者に対する安全性の相関関係を明らかにする)
- ▶ どのような加工がID識別, トレース推定に強いのか?
→ 参加者の皆さんで決める! (正直, どうなるか私にも分かりません)



【超重要】エントリーは下記HPから. まだ間に合います! (エントリー×:9/3)

<http://www.iwsec.org/pws/2019/entry/entry.html>

ご清聴有難うございました