



PWS Cup 2019: ID識別・トレース推定に強い位置情報の匿名加工技術を競う

○村上 隆夫¹, 荒井 ひろみ², 井口 誠³, 小栗 秀暢⁴, 菊池 浩明⁵, 黒政 敦⁶,
中川 裕志², 中村 優一⁷, 西山 賢志郎⁸, 野島 良⁹, 波多野 卓磨¹⁰,
濱田 浩気¹¹, 山岡 裕司⁴, 山口 高康¹², 山田 明¹³, 渡辺 知恵美¹⁴

1 産業技術総合研究所

2 理化学研究所

3 Kii株式会社

4 株式会社富士通研究所

5 明治大学

6 富士通クラウドテクノロジーズ株式会社

7 早稲田大学

8 株式会社ビズリーチ

9 情報通信研究機構

10 日鉄ソリューションズ株式会社

11 NTT セキュアプラットフォーム研究所

12 株式会社NTTドコモ

13 株式会社KDDI総合研究所

14 筑波技術大学

目次

PWSCup2019のルール

結果発表

今後の展望

(注: 来年のPWSCupは未定)

PWSCup2019

特徴

- 初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(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
チーム数	13	15	14	14	21



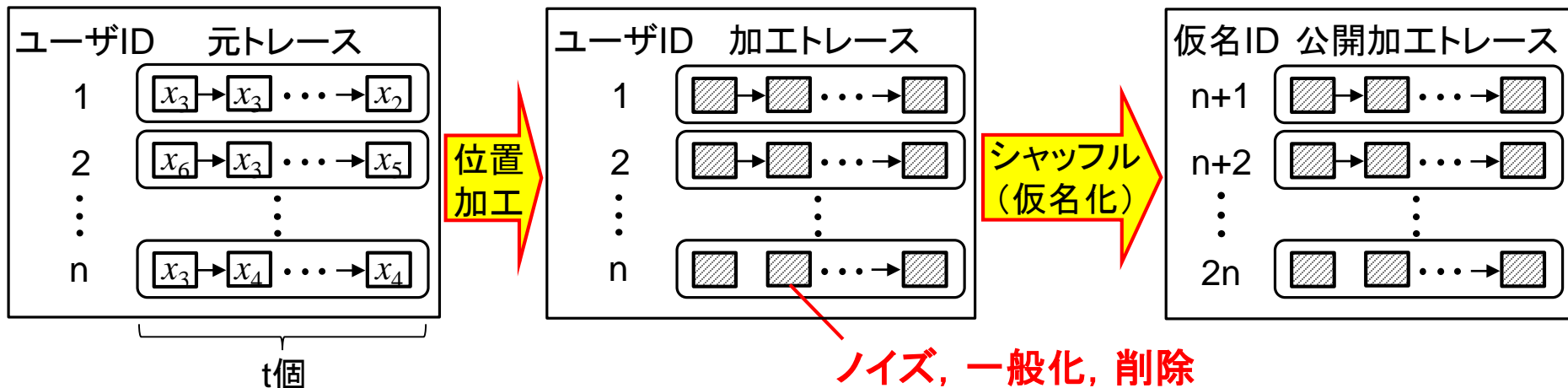
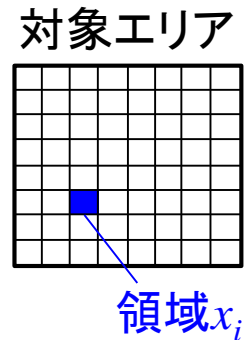
PWSCup2019: 位置情報コンテスト

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

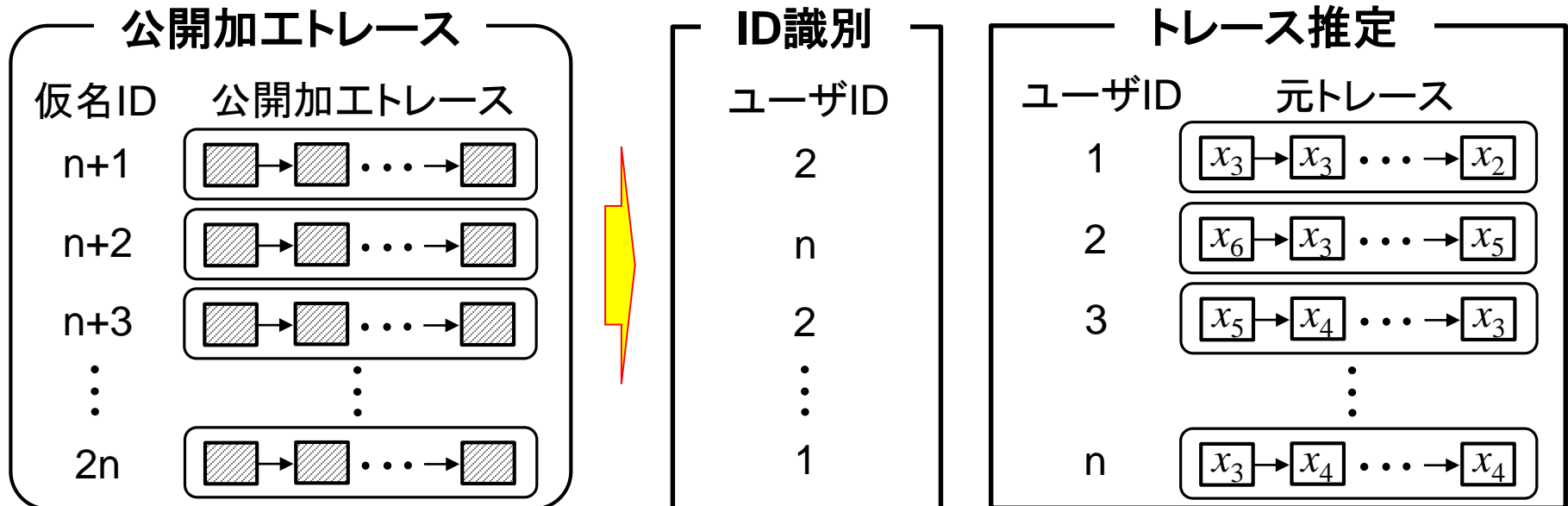
匿名加工 (Anonymization)

- 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 (例: $x_1 \Rightarrow \{x_1, x_2, x_4\}$, $x_1 \Rightarrow \{x_2, x_3, x_4\}$)
 - 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



PWSCup2019: 位置情報コンテスト

- ▶ ID識別 (ID Disclosure)
 - ▶ 各公開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 (例えば, ID識別後に位置を推定)
 - ▶ 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？

- ▶ 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

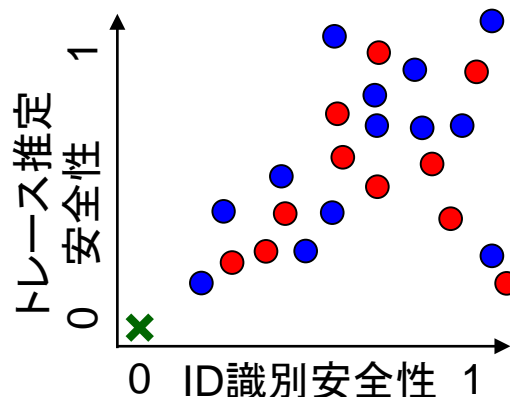
- 各チームに「ID識別対策用」、「トレース推定対策用」の2つの元トレースを配る
 - ID識別対策用 = IDP (ID Protection), トレース推定対策用 = TRP (Trace Protection)

匿名加工フェーズ:

- 元トレース(最大2個)の位置を加工し, 提出する(運営側でシャッフル)
- IDPのID識別安全性, TRPのトレース推定安全性を基に匿名加工の賞を決定

ID識別・トレース推定フェーズ:

- 他チームのIDPとTRP+実行委員の仮名化トレースをID識別・トレース推定する



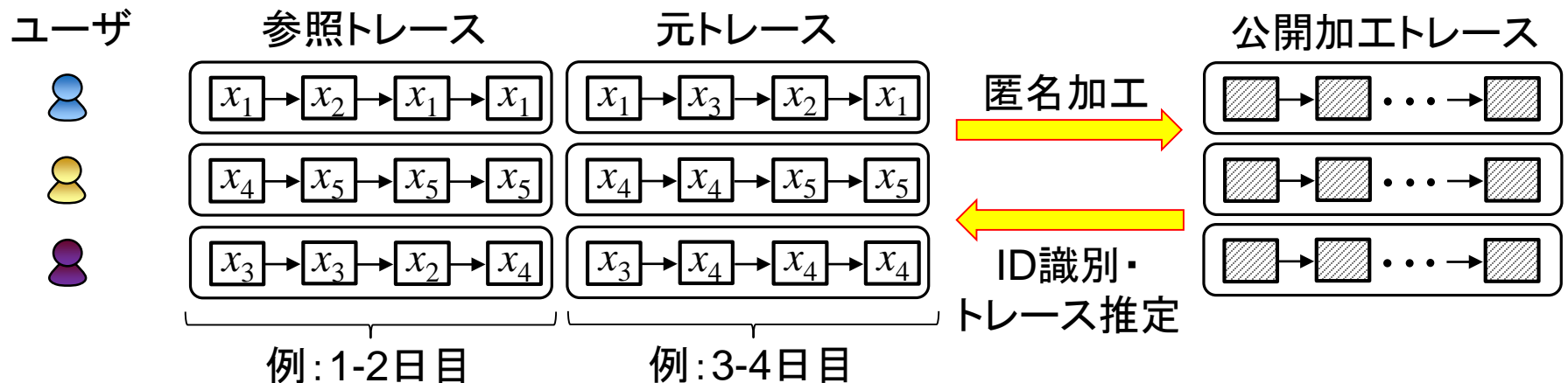
- IDP (ID識別対策用)
- TRP (トレース推定対策用)
- ✕ 仮名化トレース

〔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〕

スコア: 0(悪い) - 1(良い)

最大知識モデル or 部分知識モデル？

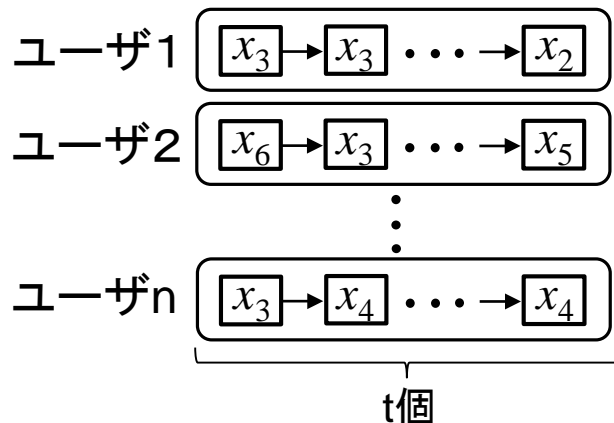
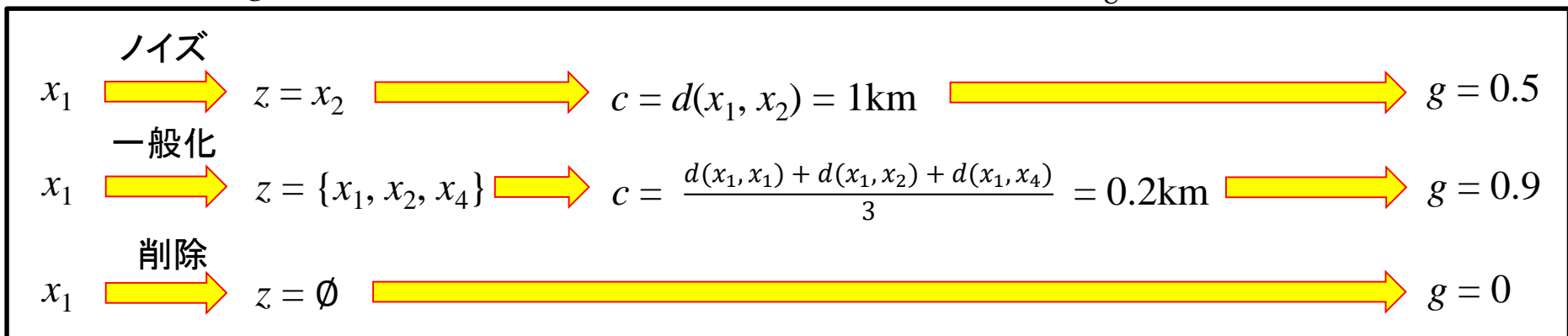
- ▶ 最大知識モデル[Domingo-Ferrer+, PST15]
 - ▶ 攻撃者が元データを知っているというモデル(攻撃者 = 提供元事業者)
 - ▶ 【課題】位置情報は特異性が高く[Montjoye+, SR13], すぐID識別される
- ▶ 部分知識モデル(PWSCup2019)
 - ▶ 攻撃者が元データを知らないというモデル(攻撃者 = 提供先事業者)
 - ▶ 攻撃者は参照トレースを入手. これを手掛かりに, ID識別・トレース推定する
 - ▶ このモデルでのコンテストになるよう, PWSCup2019用人工データを生成 (データセットの詳細はPWSCup HP)



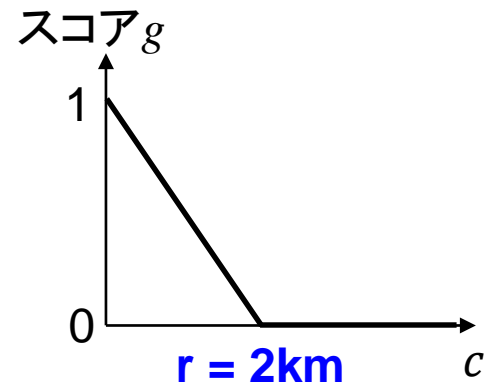
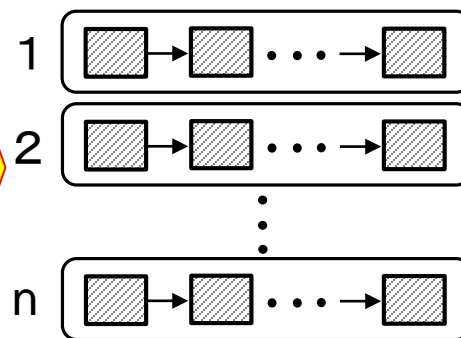
有用性指標

▶ 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(要求値:0.7)



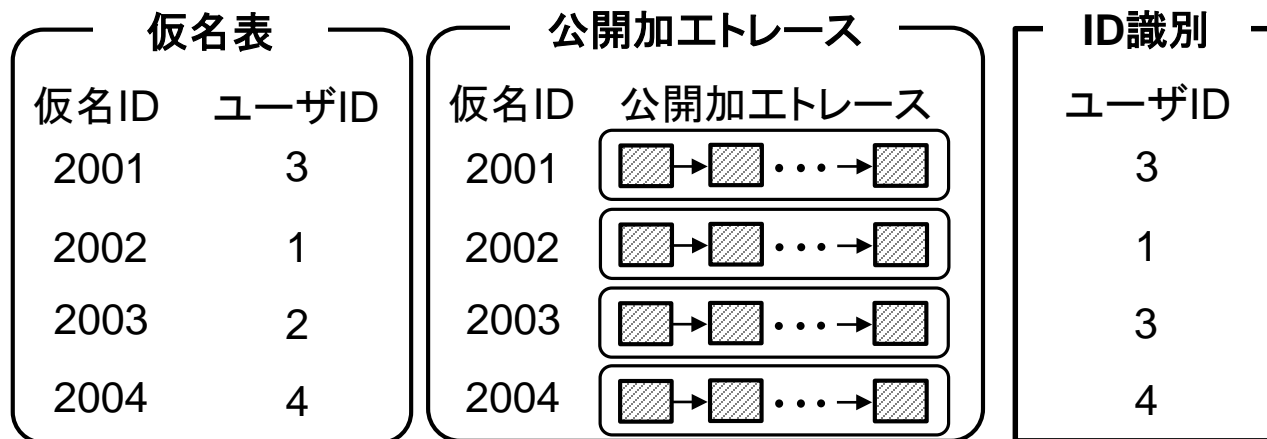
位置加工



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)



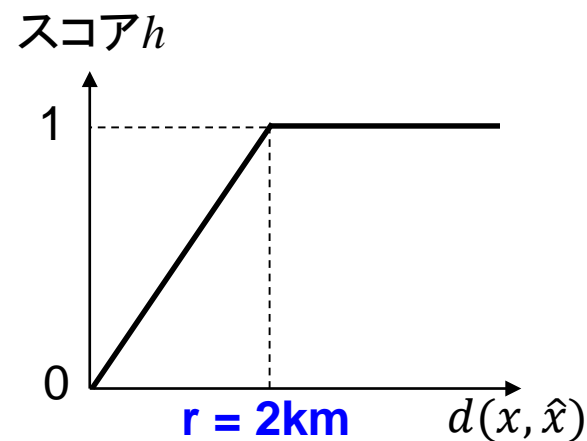
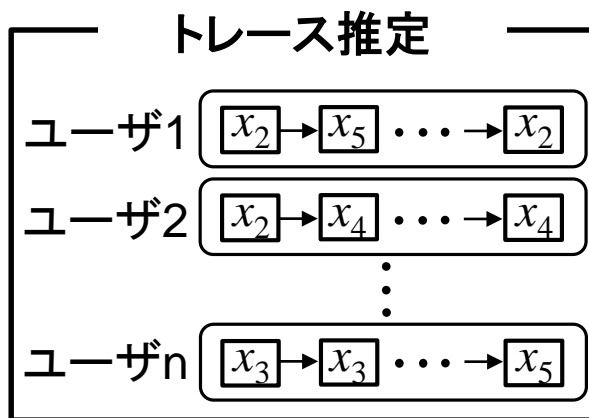
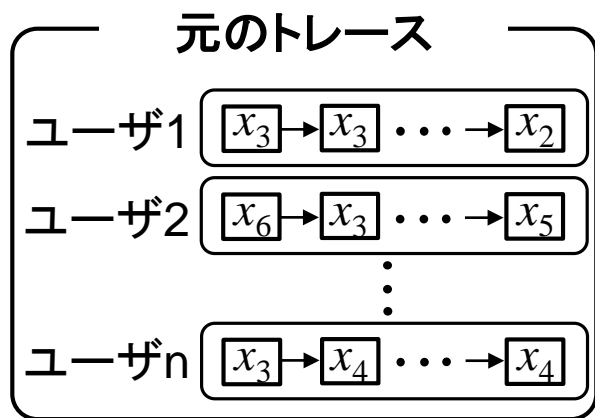
ID識別率 = $3/4 = 0.75$

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

目次

PWSCup2019のルール

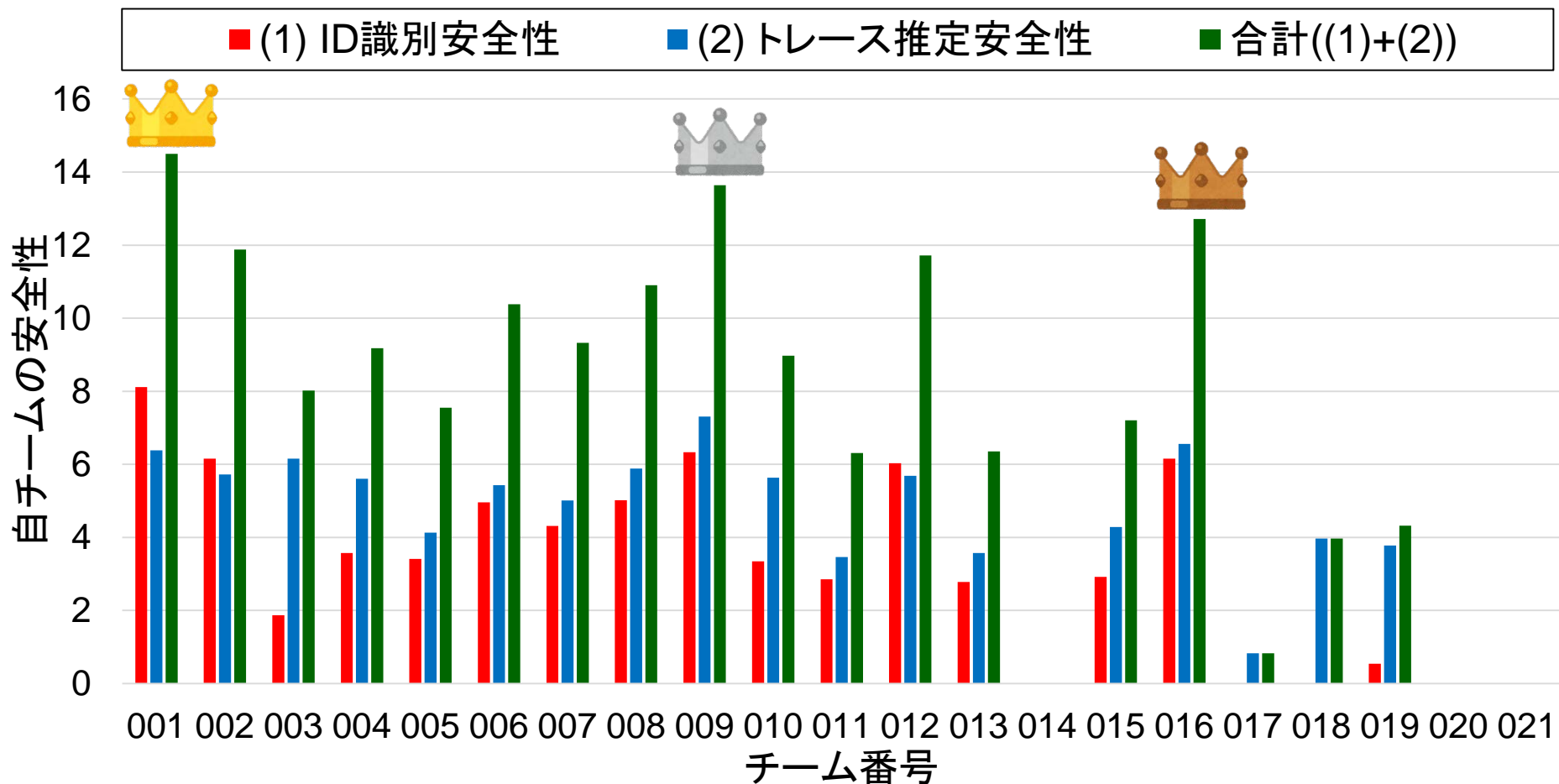
結果発表

今後の展望

(注: 来年のPWSCupは未定)

PWSCup2019の結果

- ▶ 予備戦と本戦の合計値(1:9の割合で合計)
 - ▶ 総合1位⇔2位⇔3位⇔4位以下の間には, それぞれ大きな差があり
 - ▶ (同一手法を複数データに適用したときの標準偏差 σ より遥かに大きい)

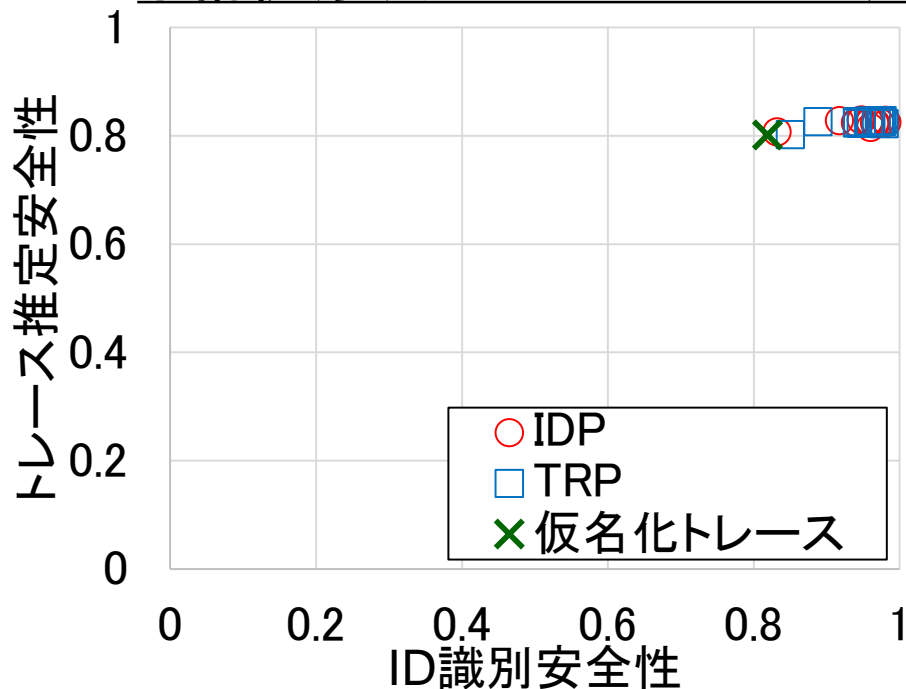


結果の詳細(2軸での評価)

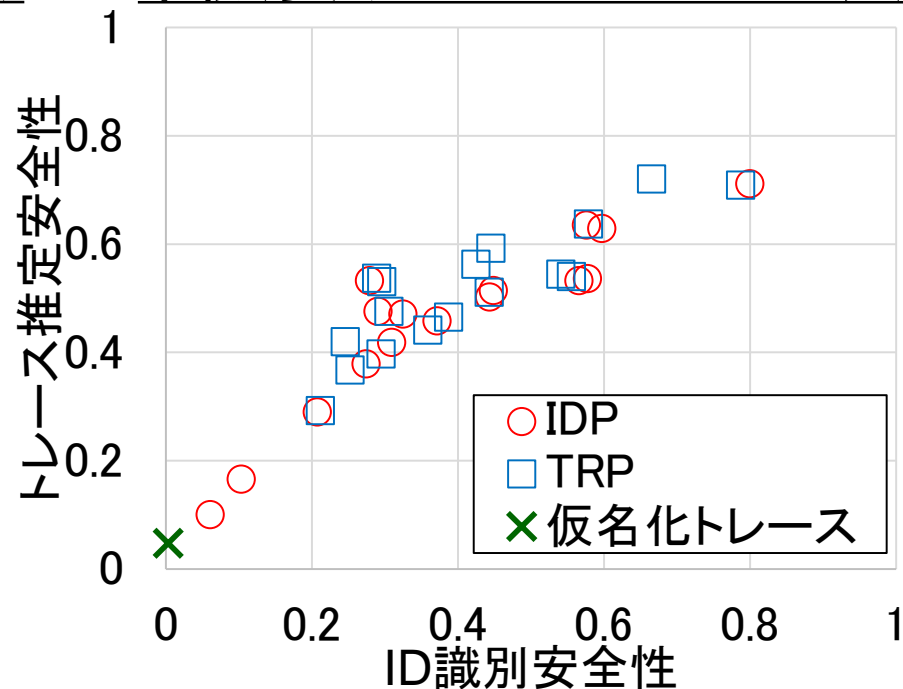
予備戦

- ▶ ID識別: トレースの長さが短く, ID識別率(= 1 - ID識別安全性)が低い
- ▶ トレース推定: 匿名加工トレースを見ずに, 参照トレースをそのまま元トレースの推定値とする「**参照トレース再送攻撃**」(or 変形版)が猛威を振るった

予備戦(参照・元トレースとも2日分)



本戦(参照・元トレースとも20日分)

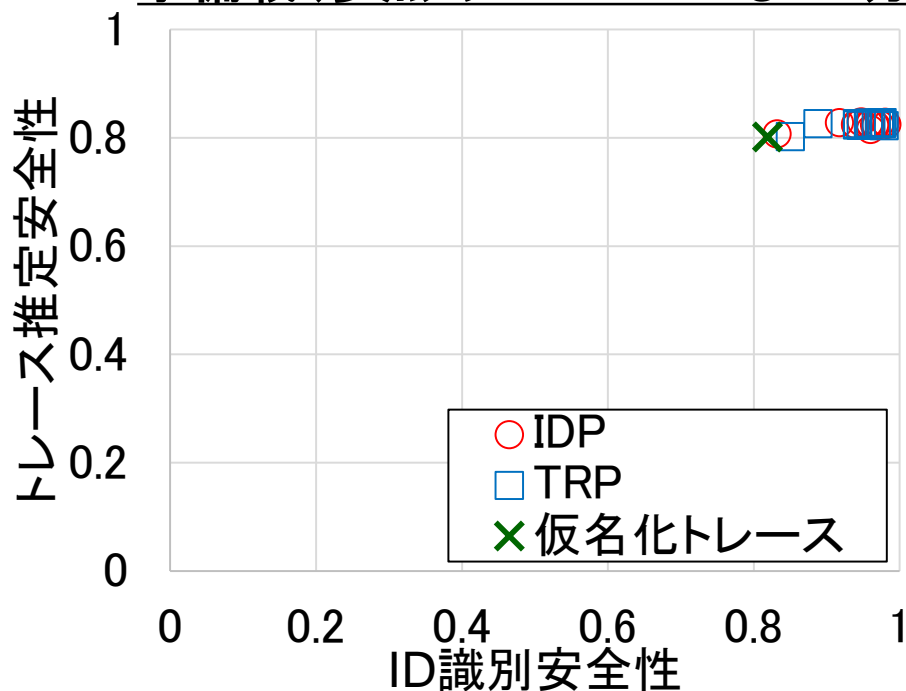


結果の詳細(2軸での評価)

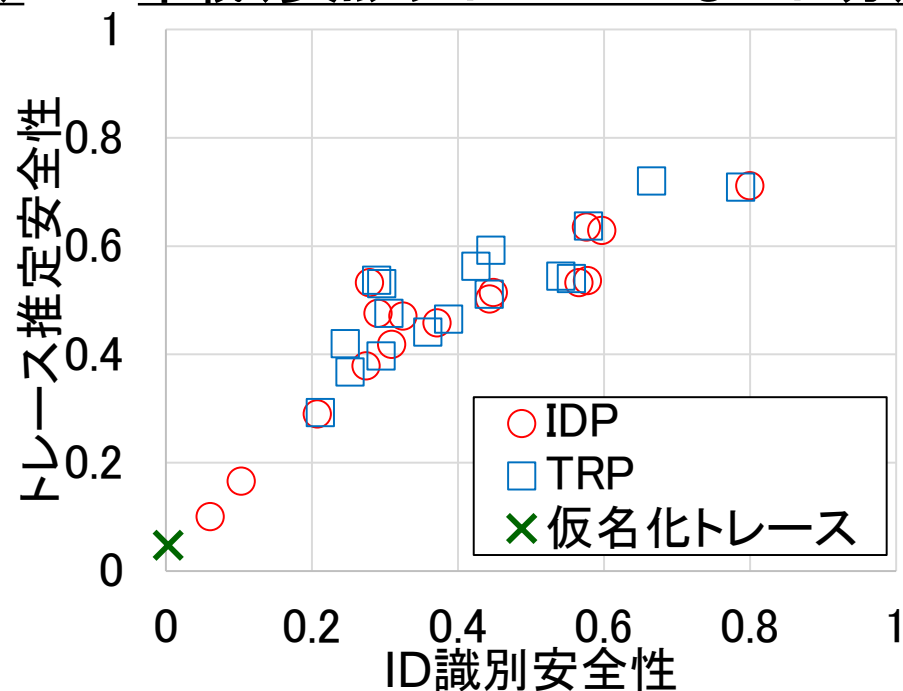
▶ 本戦

- ▶ 予備戦と比べて安全性が低下(仮名化に対するID識別率は最大**99.7%**)
- ▶ 安全性のバラつきが非常に大きい(例:ID識別安全性:0.06~0.8)
- ▶ ID識別をしてから位置情報を推定するトレース推定が有効

予備戦(参照・元トレースとも2日分)



本戦(参照・元トレースとも20日分)



目次

PWSCup2019のルール

結果発表

今後の展望

(注: 来年のPWSCupは未定)

今後の展望

▶ 有用性指標

- ▶ 今回はノイズ付与前後のユークリッド距離に基づく汎用的な有用性指標を用いたが、**アプリケーション寄りの有用性指標**を用いる方向性が考えられる
- ▶ 例1: 人気スポットの分析 → 人口分布の推定精度
- ▶ 例2: POIのタグ付け(飲食店, ホテルなど) → 訪問回数の分布[Ye+, KDD11]

▶ 個人的な見解

- ▶ データ解析系の有用性指標(例: 人口分布の推定精度)では、**ユーザ単位のスワッピング(山岡匿名化)**が有効な場合が多い
- ▶ → ID識別とトレース推定の安全性の相関が、今回の結果とは大幅に変わる

今後の展望

▶ 安全性指標

- ▶ 今年は簡単のため、平均的な安全性指標を導入
 - ▶ 例: ID識別安全性 = $1 - \text{ID識別率}$
- ▶ 課題: ID識別率を0%にすることはできない(全部1にすれば必ず1つあたる)
- ▶ → 仮に識別されても「自分でない」と否認できる指標を導入する方向性

▶ 個人的な見解

- ▶ 去年の犠牲者ゼロ精神に基づく(全ての人が十分に特定されにくい)安全性指標が一つの解? 但し, 複雑(トレース推定版を考えるとなおさら) & 自己流
- ▶ → 安全性指標/他のところを単純化する(例: ID識別だけ考える)
or 他の安全性指標(例: plausible deniability)との関係の明確化

ご清聴有難うございました

補足資料

補足資料: 有用性指標

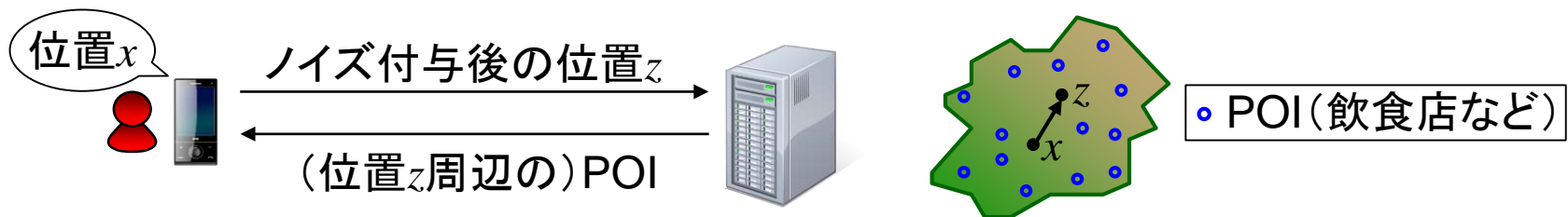
▶ 有用性の特徴

▶ LBS (Location-based Service) としての解釈:

- ▶ ユーザが (LBSプロバイダーを信用せず) 自身の位置 x にノイズを付与し, ノイズ付き位置 z 周辺のPOIを検索するアプリケーションが考えられる
- ▶ このときのPOI検索精度は, $d(x, z)$ が大きいほど低くなり, $d(x, z)$ が一定以上になると全く検索できなくなる → 有用性は, POI検索精度と相関が高い

▶ データ解析としての解釈:

- ▶ 人口分布の可視化, 人気スポットの分析, POIカテゴリーの自動タグ付けなど, 様々なアプリケーション寄りの有用性との相関が高いと考えられる
- ▶ 詳細な分析実験は今後の課題



補足資料: 有用性指標

▶ 有用性指標とPOI検索精度の相関分析

▶ データセットと匿名加工アルゴリズムは、以下のものを使用

▶ データセット:

- 疑似人流データから抽出したトレース
- ユーザ数: 1000人, 領域数: 32×32 , 時間間隔: 30分以上, 1ユーザあたりの位置: 35個

▶ 匿名加工アルゴリズム(計29個):

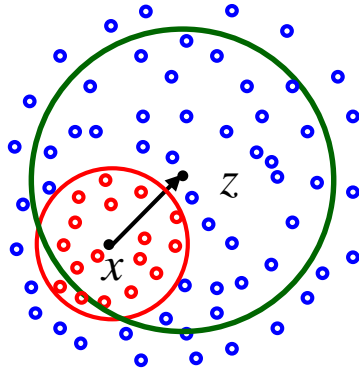
- A1-none,
- A2-MRLH ((0, 0, 0.5), (0, 0, 0.8), (0, 0, 0.95), (1, 1, 0), (1, 1, 0.5), (1, 1, 0.8), (2, 2, 0), (2, 2, 0.5), (2, 2, 0.8))
- A3-kRR (0.01, 0.1, 1, 2, 3, 4, 5, 6, 6.93)
- A4-PL: ((0.01, 1), (0.1, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (6.93, 1))
- A5-YA

▶ POIとしては、疑似人流データから東京(緯度: 35.65-35.75, 経度: 139.68-139.8)の「food」カテゴリー(例: レストラン)のPOIを全て抽出 → 計4692個

▶ 真の位置 x からの検索半径 $s_1 = 0.5\text{km}$

▶ ノイズ付き位置 z からの検索半径 $s_2 = 0.5\text{km}, 1\text{km}, 1.5\text{km}, 2\text{km}$

($s_2 = 1\text{km}, 1.5\text{km}, 2\text{km}$ に対しては、それぞれ約4, 9, 16倍の通信量を許容する)



- ユーザが検索したいPOI
- それ以外のPOI
- 真の位置 x から半径 $s_1 = 0.5\text{km}$ の円
- ノイズ付き位置 z から半径 $s_2 = 1\text{km}$ の円

補足資料: 有用性指標

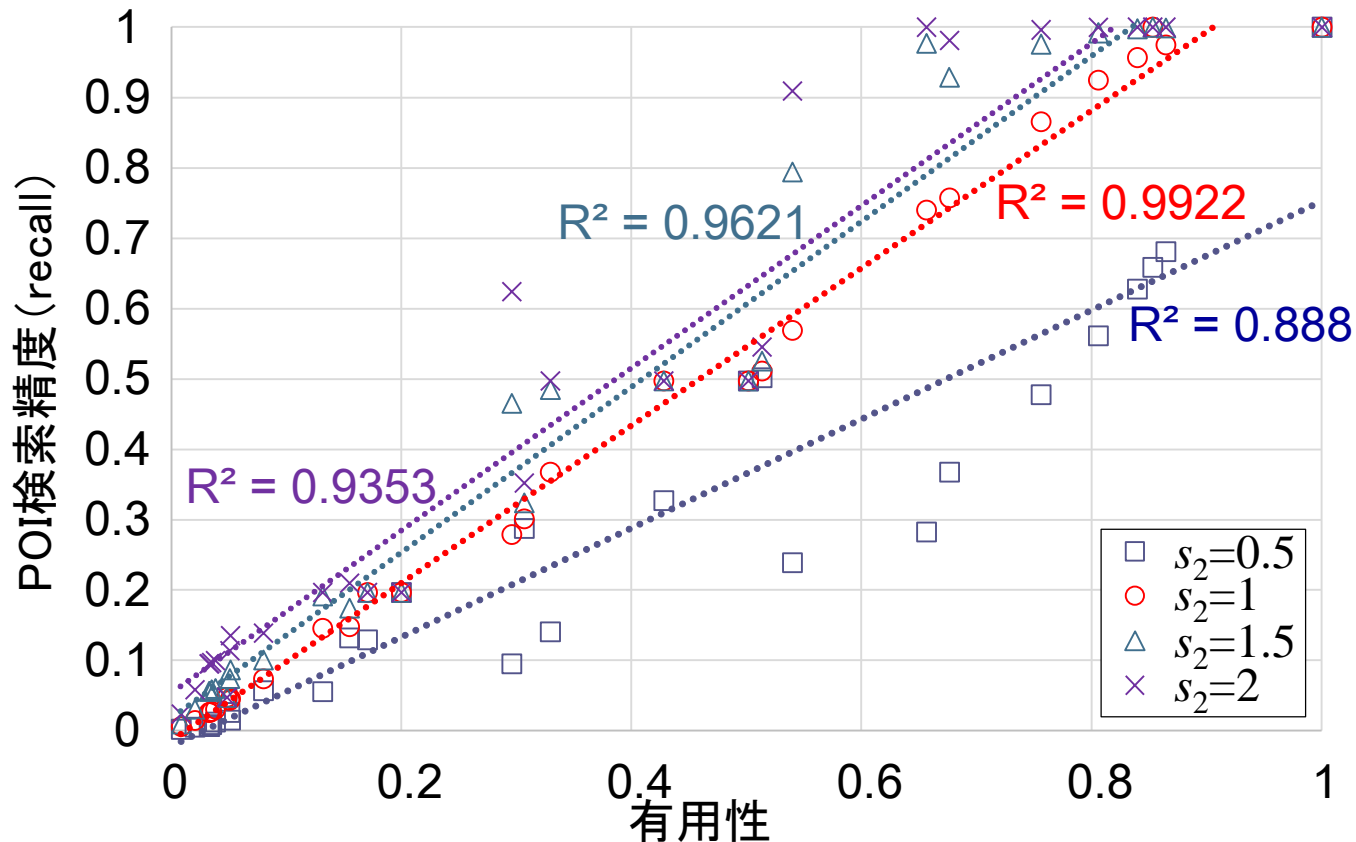
▶ 実験結果

- ▶ 検索半径 $s_2 = 1\text{km}$ のとき, 相関が非常に高い($R^2 = 0.9922$)



▶ 有用性指標のパラメータ

- ▶ 有用性の要求値は0.7 → (検索半径を2倍にしたときの)POI精度: 80%






補足資料: 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
	$x_5 \rightarrow x_5 \rightarrow x_4 \rightarrow x_3$	$x_5 \rightarrow x_4 \rightarrow x_3 \rightarrow x_3$
	$x_2 \rightarrow x_2 \rightarrow x_4 \rightarrow x_3$	$x_2 \rightarrow x_3 \rightarrow x_2 \rightarrow x_4$
	$x_1 \rightarrow x_4 \rightarrow x_4 \rightarrow x_1$	$x_1 \rightarrow x_1 \rightarrow x_4 \rightarrow x_4$

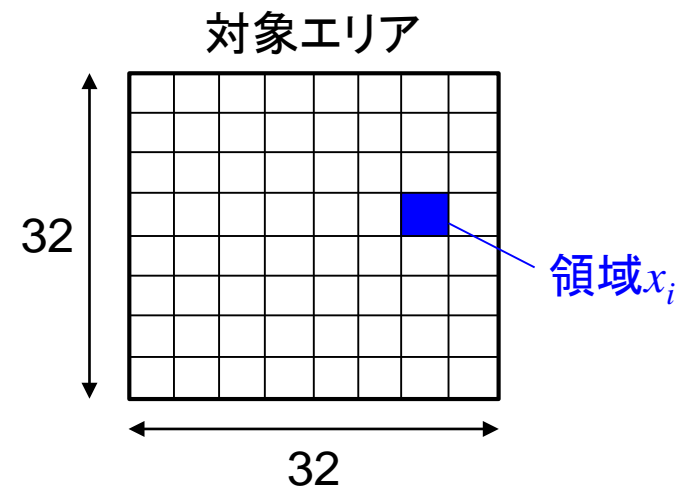
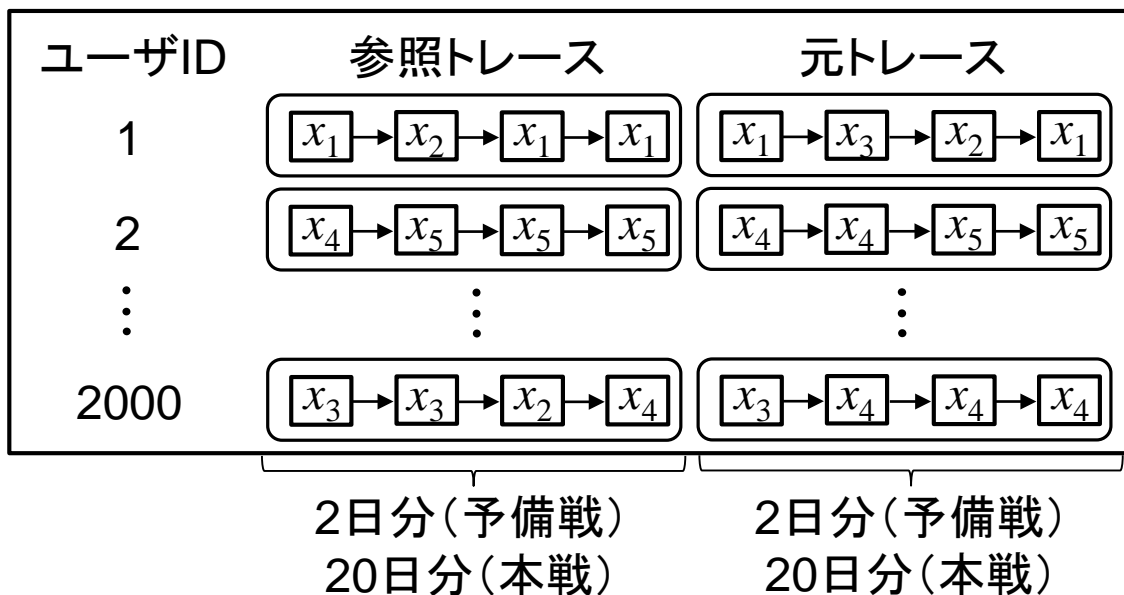
疑似人流
データ



ランダムな
生成モデル

補足資料: データセット

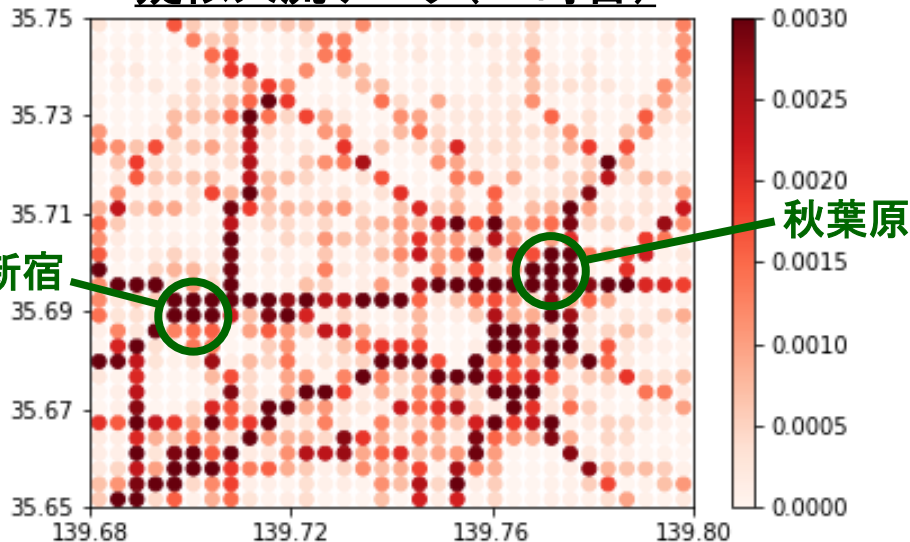
- ▶ PWSCup2019用人工データ(詳細)
 - ▶ ユーザ数: $n = 2000$
 - ▶ 位置情報数: $m = 1024$ (東京中心部を 32×32 の領域に分割)
 - ▶ トレースの長さ: 1日あたり20個の位置情報(8~18時, 30分おき)
 - ▶ 予備戦: 参照トレース・元トレースともに2日分
 - ▶ 本戦: 参照トレース・元トレースともに20日分



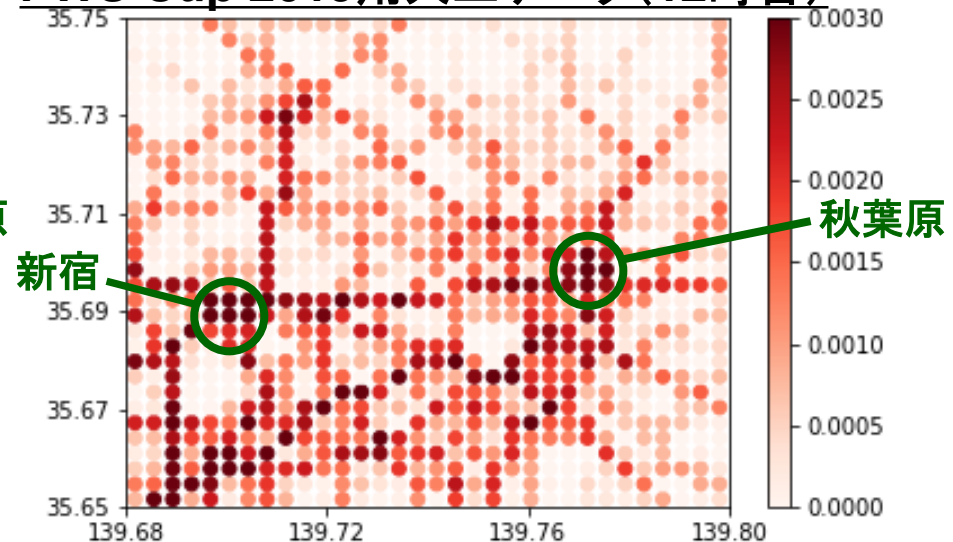
補足資料: データセット

- ▶ 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: 1024x1024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

疑似人流データ(12時台)



PWS Cup 2019用人工データ(12時台)



補足資料: 公平性に関して

▶ 公平性の課題

- ▶ 各チームに配布するデータセットが異なる
- ▶ 特に, 病院領域(通称:ドラ)の割合がチーム毎に異なると不公平...



▶ 今回の対応

- ▶ (1) **ドラの割合が全チームで同じ**になるように人工データを生成
 - ▶ 受賞チームは, 予備戦と本戦の安全性を1:9の割合で合計して決める
 - ▶ → 本戦データを大量に(200チーム分)生成し, そこから各チームのデータを,
▶ 「**予備戦のドラ割合 + 9 * 本戦のドラ割合 = 0.0445**」になるように選定
- ▶ (2) これ以上のことは, 「コンテストなので割り切る」
 - ▶ とは言え, 今回の本戦結果は, かなりアルゴリズムによる差が出ています