## PWS Cup 2024 Team20: KAT-TUNE

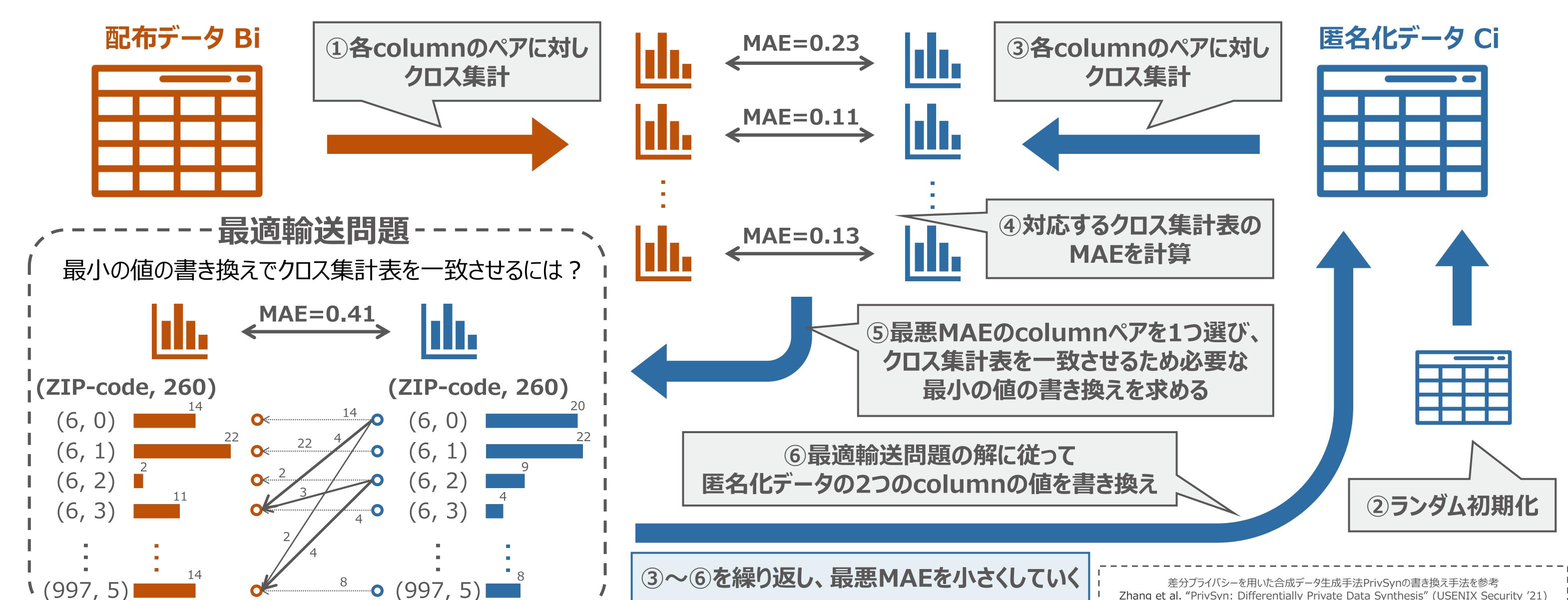
奥井 株式会社 KDDI総合研究所 福嶋、日比、飯澤、秋元、湯野川 株式会社 ARISE analytics

椎名、篠田、中谷、吉田 トヨタ自動車 株式会社

三社の調和

-TUNE

### 匿名化手法:最適輸送を用いた合成データ生成



### 攻撃手法: ハンガリアンアルゴリズムを用いた全体最適攻撃

個人特定攻擊

DB再構築攻撃

攻撃フェーズ の目標

Ratingsデータの各レコード(以下 $B_i$ )に紐づく基本属性データのレコード (以下 $A_i$ )を推定すること。

仮説

予測精

度の評

価

影響度

合いの

算出

 $B_i$ のペアとなる $A_i$ を被りなく紐づける組み合わせの中では、正しいペアの 組合せを選んだ際に、全ペアの紐づく確率は最大化されるのではない か?

Ratingsデータで黒塗りされた元データの値を推定すること。

正しいペアが選択できているのであれば、黒塗り部分に正しいRatingの 値を代入した際に、そのペアの紐づく確率は最大化されるのではないか?

### ①匿名加工データから同時確率を算出

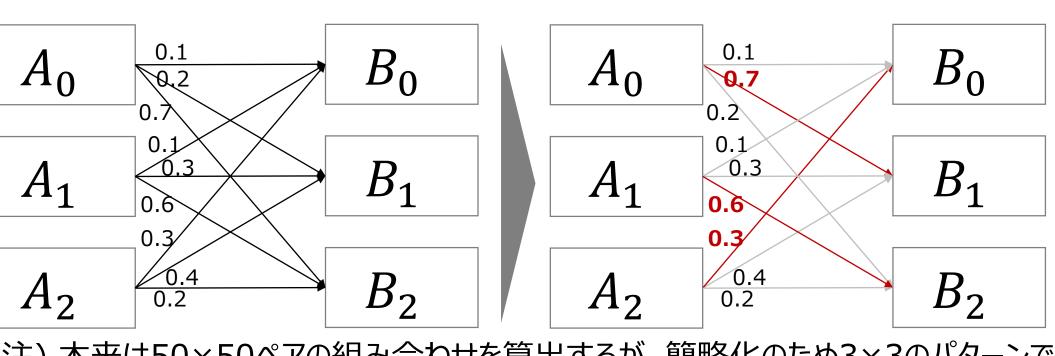
$$P(A_i, B_j) = P(X_0, \dots, X_{49}) = \prod_{k=0}^{49} P(X_k | X_{k<})$$

$$= \prod_{k=0}^{49} \left( P(X_k) \prod_{l=0}^{k-1} P(X_k | X_l) / P(X_k) \right)$$

注) $P(X_k)$ 及び $P(X_k|X_l)$ はクロス集計表から算出可能.

匿名加工データから、ペア $(A_i, B_i)$ が紐づく確率として、同 時確率 $P(A_i, B_i)$ を算出。

### ②被りのないペアの組み合わせを算出



注) 本来は50×50ペアの組み合わせを算出するが、簡略化のため3×3のパターンで例示

ハンガリアンアルゴリズムを用い、全ペアの紐づく確率(総 積)が最大となる"ペアの組合せ"を算出。

#### ③黒塗りされた値を推定

Zhang et al. "PrivSyn: Differentially Private Data Synthesis" (USENIX Security '21) ※ 今回の匿名化には差分プライバシーは適用していない)

> $\hat{r} = arg \max P(A_i, B_{ii} | *= r)$  $r \in \{0,1,2,3,4,5\}$

注)  $j_i$ は個人特定攻撃によって推定されたiのペア

同時確率が最大となるRating値を算出

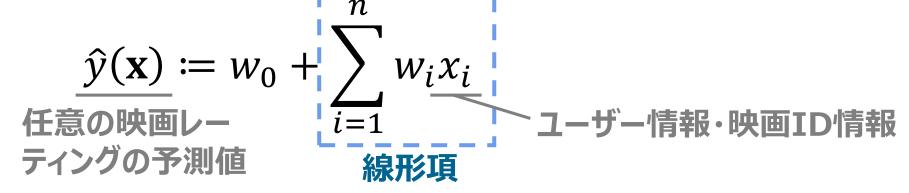
### ■データサイエンス:推薦システム開発におけるKAT-TUNE匿名化手法の有用性評価

問い:匿名化データは推薦システム開発に有用なのか? → 手法:匿名化後でも特徴量間の相互作用効果が残っているかを評価することで検証

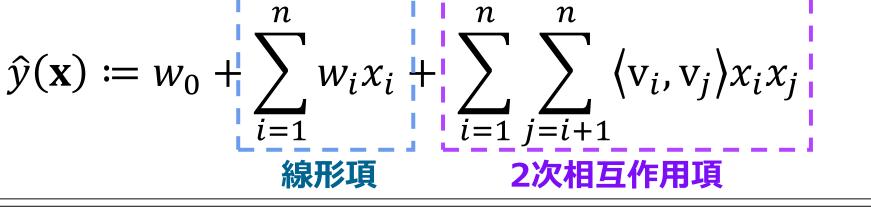
#### 2次相互作用の影響度合いの評価手法

任意のユーザーに対する、任意の映画のレーティング値を予測する モデルの精度を評価する。

#### i)相互作用を考慮しないモデル [線形項のみ]



ii)相互作用を考慮したモデル(FM)\*1 [線形項+2次相互作用項]



i)とii)を比較して2次相互作用項の影響を評価する

2次相互作用の 影響度合い

ii ) での予測精度

i )での予測精度

匿名化前後で

変わらないと良い

point

# 匿名化前の影響度合い 推薦システム 開発に有用と考 KAT-TUNE えられる匿名化 本コンテストの有用性スコア

匿名化前の影響度合い 参考) ランダムシャツ フルによる匿名化 -0.010 本コンテストの有用性スコア

凡例: 〇 イテレーション数の異なる匿名化

### 【結果·示唆】

KAT-TUNE手法だと相互作用の影響度が低くなっている

- ・推薦システム開発で間違った示唆を与える可能性
- ・相互作用も評価に入れた上での手法開発が必要と思慮
- \*1 Factorization Machines, Rendle, S. (2010), Factorization Machines, WWW