

Algorithm 1 結合処理

Input: 2 値データセット \mathcal{T}_b , 多値データセット \mathcal{T}_p , 匿名化パラメータ k , アイテム数 m

Output: k^m -匿名性を満たした混合型データセット \mathcal{T}'

- 1: $A = \mathcal{T}_b$ の列数, $V = \mathcal{T}_b$ の列番号の組からなる部分集合 ($|V| = AC_{m-1}$)
- 2: $\mathcal{T}' = \mathcal{T}_b$
- 3: **for** i **in** \mathcal{T}_p の列数 **do**
- 4: \mathcal{T}' の最終列に \mathcal{T}_p の i 列目を追加
- 5: **for** v **in** V **do**
- 6: \mathcal{T}' の v 列目と $A+i$ 列目で構成されるデータ τ を作成
- 7: **if** τ が k^m -匿名性を満たさない **then**
- 8: τ が k^m -匿名性を満たすよう \mathcal{T}' の $A+i$ 列目を一般化
- 9: **return** \mathcal{T}'

手法を用いて, 匿名化を行う. その後, 今回, 新たに提案する k^m -匿名性を満たすための結合処理を行う (Algorithm 1).

4 実験

人工的に作成した 1000×30 の混合型データで数値実験を行い, 情報損失の量を検証する.

4.1 実験設定

混合型データを 20 列が 2 値データ, 10 列が多値 (7 値) データで構成し匿名化パラメータ k は 10, アイテム数 m は 2 とする. また, 評価指標としては匿名化前後においてどの程度値が変更されたかを測るために情報損失として以下の式を用いる. このとき, N をデータ行数, M をデータ列数, ij をそれぞれ行番号, 列番号, n を多値数, $\#X$ を値の要素数とする.

$$loss := \frac{1}{NM} \sum_{ij} \begin{cases} |X_{ij} - X'_{ij}| & \text{2 値データ} \\ \frac{\#X'_{ij} - \#X_{ij}}{n} & \text{多値データ} \end{cases}$$

実験は 25 回ずつ行ない, その平均値 ($loss_ave$), 最大値 ($loss_max$), 最小値 ($loss_min$) を求める.

4.2 実験結果

実験結果を表 1 に示す. ここで, \mathcal{T}'_b , \mathcal{T}'_p , \mathcal{T}' における値はそれぞれ M が, 20, 10, 30 のときのものであることに注意されたい.

表 1 情報損失量の評価

	\mathcal{T}'_b	\mathcal{T}'_p	\mathcal{T}'
$loss_ave$	0.00117	0.04713	0.01649
$loss_max$	0.00125	0.04713	0.01654
$loss_min$	0.00105	0.04713	0.01641

5 考察

混合型データを分割した後, それぞれに匿名化処理を行なう手法であるため, 情報損失量が大きくなると思われていた. しかし, 図 1 を見ると, 全体の 1.7% 程度しか損失がなく, その最大, 最小の幅が小さいこともわかる. また, 結合処理では, 匿名化データの多値部分に対してさらに一般化を用いているが, 一般化は階層構造であるため, 元データに対する匿名化処理として問題がないと言える.

6 まとめ

本稿では, 混合型データに対する k^m -匿名化手法を提案した. 加えて, 数値実験により, 少ない情報損失量で, 匿名化が行えることを確かめた. 今後の課題として, 匿名化パラメータやデータ列数などの値を変更し, 本手法で今回使用したデータセットと異なる内容や形式のデータセットを用いても汎用的に使用出来るかどうか確認することが挙げられる.

参考文献

- [1] M. Kobayashi, et al. Extended k^m -anonymity for randomization applied to binary data. In *PST2023*, pp. 221–227. IEEE, 2023.
- [2] G. Poulis, et al. Apriori-based algorithms for k^m -anonymizing trajectory data. *TDP*, 7(2):165–194, 2014.