

An Efficient Local Recoding Anonymization in Data without Hierarchical Taxonomies

階層的分類を伴わない効果的な匿名化のためのローカルリコーディング

Mohammad Rasool Sarrafi Aghdam

Professor Noboru Sonehara

どんな研究？

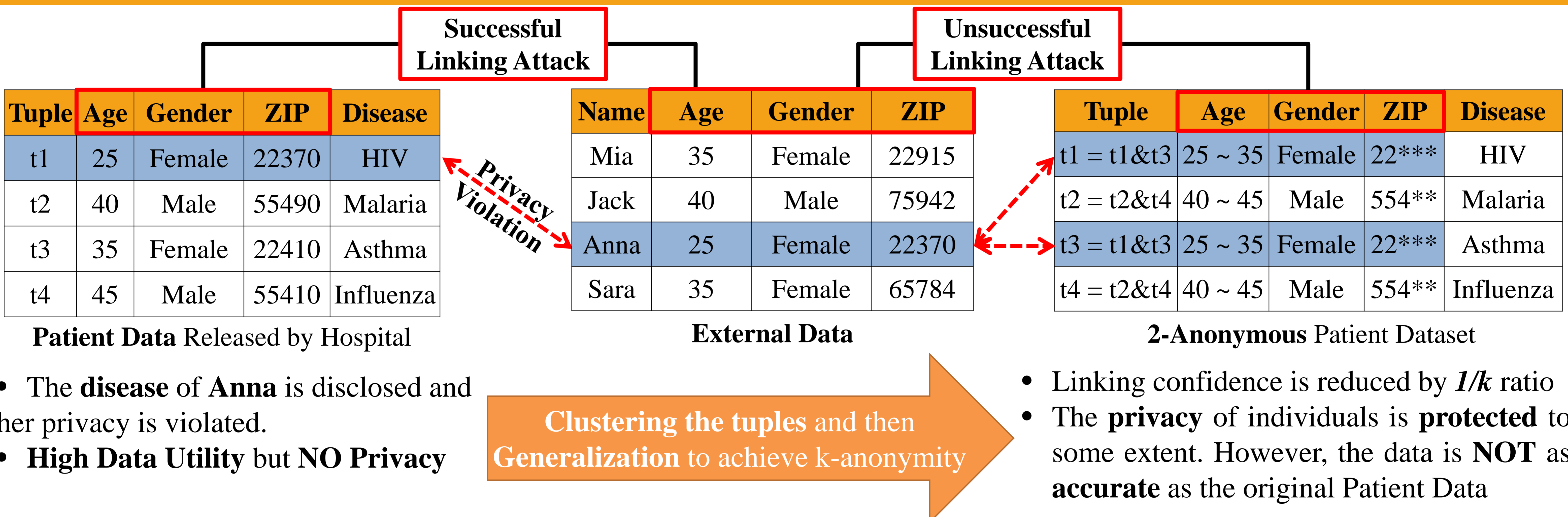
個人情報を保護するために、データを公開する際にプライバシー保護のために広く用いられているモデルとしてk-匿名性(k-anonymity)がある。

k-匿名性とは、個人情報から個人を特定することを難しくするための技術である。しかしながら、k-匿名性を用いたデータは情報が喪失されるため正確さに欠ける。現存するk-匿名化アプローチは大規模な情報喪失という点において問題がある。そこで我々はattributes hierarchical taxonomies から独立した数値属性とカテゴリ属性を含むデータ間の距離の計算にもとづく新しいモデルとSpatialDistance (SD) ヒューリスティックアルゴリズムを提唱する。我々の研究では、SDは既存のよく知られているアルゴリズムに比べて有意に情報喪失を減少させることが示されている。

Abstract

To protect the privacy of individuals, a model that is widely used for privacy preservation in publishing micro-data, is **k-anonymity**. It reduces the linking confidence between sensitive information and specific individual. However, k-anonymous dataset loses its accuracy due to the **information loss**. Most of the existing k-anonymization approaches suffer from huge information loss. We propose a new model and SpatialDistance (SD) heuristic algorithm based on **distance calculation** between tuples including **numerical** and **categorical** attributes which is independent of attributes hierarchical taxonomies and **enhances the data utility** significantly.

状況設定



研究内容

SD algorithm uses distance calculation for clustering the tuples for generalization through local recoding to Enhance the Utility of Data in k-anonymization.

$$\text{Total Distance } (t_i, t_j) = D(\text{Numerical}(t_i, t_j)) + D(\text{Categorical}(t_i, t_j))$$

- Numerical Attribute:** $D(t_j, t_k)_{Att_i} = \frac{|x_j - x_k|}{R(Att_i)}$
- Categorical Attribute:**
 - Cardinality (Att_i) = 1 then Distance (a,a) = 0
 - Cardinality (Att_i) = 2 then Distance (a,b) = 1
 - Cardinality (Att_i) > 2
 - Measuring similarity based on observation probability
 - Defining distance

	Japan	USA	China
Male	5	5	1
Female	3	0	6

- Assuming, values in tuple 1 are **Male** (Gender), **Japan** (Nationality)
- Japan is closer to USA than China**
 - Because the probability of being Male from Japan and Male from USA is closer

