個人特定のリスクを低減させる匿名化技術

千田　浩司

匿名化は，プライバシーを守りつつ個人のデータを他者に提供する手段として従来用いられている．しかし氏名などを削除するだけの古典的な匿名化は，たとえばインターネット上で公開されている各種データとの突き合わせにより，個人を特定されうることが問題視されている．わが国における個人情報保護法の改正においても，個人を特定できないように個人情報を加工する「匿名加工」の方法を施行までに定めるとしており，その内容が待たれるところである．本稿では，古典的な匿名化では守れない個人特定やプライバシー侵害のリスクについて触れ，特に個人特定のリスクを低減させる匿名化技術とその課題について概説する．

キーワード：匿名化，プライバシー，パーソナルデータ

1. はじめに

ICT (Information and Communication Technology) の発達に伴い，多種多様な大量の情報が容易に収集できるようになり，情報の利活用による新たな価値創造への期待が高まっている．われわれのパーソナルデータ（個人に関するデータ），たとえば買物履歴，インターネットアクセス履歴，位置・移動情報，バイタル情報も事業者などによって日々蓄積され，研究やサービス向上などのための分析素材として利活用され始めている．しかしパーソナルデータを扱う際は個人情報やプライバシーの保護に十分配慮する必要がある．

プライバシーを守りつつパーソナルデータを利活用，特にパーソナルデータを蓄積する機関とは異なる機関（以降，第三者とよぶ）へ提供する有効な手段として，匿名化が挙げられる．匿名化は，誰のパーソナルデータか特定できないようパーソナルデータを加工する処理を指し，その方法はさまざまである．すなわち匿名化にはそれぞれ異なる特徴があり，状況に応じて適切な方法を選択することが望ましい．

氏名などの削除は古典的な匿名化としてよく知られるが，それだけでは不十分な場合がある．たとえば時刻と詳細な位置情報（緯度経度情報）を含むパーソナルデータは，夜間には自宅にいる可能性が高いことから自宅の場所が推測され，個人特定につながるかもしれない．また有名な事例として，氏名を削除した患者の医療情報（診断結果，投薬情報などに加え性別，生年月日，ZIP コードが含まれていた）を米国マサチューセッツ州が公開したところ，マサチューセッツ州知事の医療情報が特定されてしまった [1]．同じく公開されている投票者名簿（氏名，性別，生年月日，ZIP コードが含まれていた）から，州知事と同じ性別，生年月日，ZIP コードの人がいないことが明らかになったためである．さらには，1990 年の米国国勢調査の回答データは性別，生年月日，および 5 桁の ZIP コードだけで全体の 87%が一意の情報になっているという報告もある [2]．このように，情報の一意性やほかのパーソナルデータとの突き合わせなどにより，氏名などを削除しただけの古典的な匿名化ではプライバシーが損なわれるリスクがある．一方で，非常に強力な匿名化を施したパーソナルデータは，利用価値を著しく損ねかねない．いわゆる匿名性と有用性のトレードオフの問題がある．

本稿では，古典的な匿名化では守れない個人特定やプライバシー侵害のリスクについて触れ，特に個人特定のリスクを低減させる匿名化技術について紹介する．また，既存の匿名化技術の課題と今後の展望について述べる．

2. パーソナルデータの開示リスクと匿名化

われわれのパーソナルデータはさまざまな形で事業者などに蓄積されている．そして蓄積されたパーソナルデータは，研究やサービス向上などのために利活用，特に第三者へ提供される場合もある．それではパーソナルデータを第三者に提供する際，どのような匿名化が適切といえるだろうか．パーソナルデータを提供している個人からすれば，第三者には自分のパーソナルデータを知られたくないと思う人も少なくないだろう．

官庁統計においては，個票データ（個人や企業などの個体による，調査票に対する個々の回答内容）の中の個体が誰であるかわかってしまうことを個体の識別と

ちだ　こうじ
日本電信電話（株）NTT セキュアプラットフォーム研究所
〒 180–8585　東京都武蔵野市緑町 3–9–11

表 1　元のパーソナルデータのテーブル

name（正識別子）	sex（準識別子）	age（準識別子）	income（センシティブ属性）	item 1（非センシティブ属性）
Alice	F	24	\$46K	coffee
Bob	M	25	\$52K	beer
Chris	M	30	\$57K	cola
Dan	M	30	\$81K	milk
Eve	F	32	\$50K	cola
Flora	F	32	\$104K	whiskey

よび，個体の識別の起こる危険性を開示リスク，あるいは識別リスクとよぶ [3]．個体が識別されることによって，個票データに含まれるセンシティブな情報が個体と紐づくことが問題となる．また開示リスクは，識別（開示）(identification disclosure) リスクおよび属性開示 (attribute disclosure) リスクに分類される [4]．後者は特定の個体のセンシティブな情報が第三者に（狭い範囲で）知られるリスクだが，識別されなくても属性開示が起こりうる（4 節で例示する）．パーソナルデータの第三者提供においては，識別が個人特定，属性開示がプライバシー侵害の問題となり，これら開示リスクを個人が許容できるレベルまで低減させることが重要といえよう．もちろん，リスクの低減は匿名化だけでなくセキュリティ対策などを組み合わせた複合的な対処が効果的となる．

本稿では，前節で述べたように個人特定のリスク，すなわち識別リスクを低減させる匿名化技術を中心に紹介する．属性開示については 4 節で触れる．

2.1　匿名化技法

本稿で対象とする元のパーソナルデータは，表 1 に例示するように各個人のデータが 1 レコードに記載されたテーブルとする．そして各列の属性は以下のいずれかに分類できるものとする．

- **正識別子**：個人を一意に識別できる属性，または当該属性の組．たとえば氏名，住所のような属性の組み合わせは無視できない確率で正識別子となる [5].
- **準識別子**：間接的に個人を識別できる属性．性別や年齢のような属性は間接的に個人の識別に用いることができる [3].
- **センシティブ属性**：正識別子，準識別子以外で，個人のプライバシーに関するものなど，他人にむやみに知られたくない属性．センシティブ属性の値をセンシティブデータとよぶ．
- **非センシティブ属性**：上記以外の属性．非センシティブ属性の値を非センシティブデータとよぶ．

表 1 に例示するようなテーブルを匿名化するために，さまざまな技法（加工方法）が知られている．表 2 は，先行文献 [6] の表 4 に基づき代表的な匿名化技法をまとめたものである．匿名化技法については詳細を記した先行文献がいくつかあるので参照されたい [9–11]．なおノイズ付加や PRAM などは属性値が確率的に変化する．このような加工方法は撹乱的 (perturbative) とよばれる．一般化や曖昧化のように確率的な要素を伴わない加工方法は非撹乱的 (non-perturbative) とよばれる．

2.2　識別リスクの評価指標

それでは元のパーソナルデータに対して匿名化技法をどのように適用すればよいだろうか．有効なアプローチとして，識別リスクの評価指標（以降，識別リスク指標とよぶ）を定義し，当該指標の基準値を満たすように匿名化技法を適用する方法が挙げられる．以下に既存の識別リスク指標をいくつか紹介する．なお正識別子は属性削除されているものとする．

- **母集団一意性**：官庁統計における代表的な識別リスク指標であり，キー変数（準識別子）の組み合わせについて母集団で個体が一意に定まるとき，その個体を母集団一意とよぶ [3]．同じキー変数の組み合わせをもつ個体が k 人いる場合には母集団 k 意とよぶ．母集団の一部からなる標本データについて個体が一意に定まるとき，その個体を標本一意とよぶ．識別リスクの評価において問題となるのは，標本においても母集団においても一意となる個体であり，そのような個体の推定方法が提案されている [12]．標本および母集団での一意性を回避または推定困難とするため，一般化や曖昧化，属性削除などが用いられる．
- ***k*-匿名性**：あるテーブルについて，すべての準識別子の値が等しいレコードの集合を準識別クラスとよび，すべての準識別クラスが k 個以上のレコードをもつとき，そのテーブルは k-匿名性を満たすという [1]．k-匿名性を満たす加工方法を k-匿名化とよび，準識別子に対して非撹乱的な手法が一般に用いられる．表 3 は local recoding およびセル削除を用いて 2-匿名化を行ったテーブルの例である．k-匿名性は標本 k 意の考え方に類似しているが，k-匿名性の関連研究はデータベースやセキュリティなどのさまざまな研究分野で発展を続けている．たとえば情報損失を最小とする k-匿名化は NP-困難であることが知られ [13]，有用性を

表 2　代表的な匿名化技法（先行文献 [6] の表 4 に基づき作成）

分類	技法	概要
属性情報の削除	属性（列）削除	正識別子など，開示すべきでない属性を削除する
	仮名化	開示すべきでない属性を符号や番号などに置き換える
属性情報の置換え	一般化	属性値を上位の値や概念に置き換える（例：「年齢」→「年代」，「キュウリ」→「野菜」） データ全体に行うものを global recoding，局所的に行うものを local recoding とよぶ 四捨五入や最も近い定数の倍数への変換等を丸め法 (rounding) とよぶ
	曖昧化	特に大きい，もしくは小さい値をまとめる （例：100 歳以上の人を「100 歳以上」とする） 大きい値のまとめを top coding，小さい値のまとめを bottom coding とよぶ
	ミクロアグリゲーション	複数のレコードをグループ化し，同じグループの値を代表値（例：平均値）に置き換える
	ノイズ付加	一定の分布に従った乱数的なノイズを加える
	スワッピング	レコード間で属性値を（確率的に）入れ替える
	PRAM	マルコフ推移確率行列に基づき，確率的に属性値を置き換える [7, 8] Post RAndomization Method の略
その他技法	レコード（行）削除	特殊な属性値をもつレコードを削除する （例：120 歳以上のレコードを削除）
	セル削除	開示すべきでない属性値を削除する
	疑似データ作成	元データと統計的に疑似させる人工的な合成データを作成する
	サンプリング	元データ全体から一定の割合・個数でランダムに抽出する
	並び替え	レコードの順序をランダムまたは値の大小順やアルファベット順などに置き換える

可能な限り損ねない k-匿名化のアルゴリズムが数多く提案されている [14]．また k-匿名性の派生指標として，次に紹介する Pk-匿名性や，4 節で触れる属性開示リスクを評価する指標などが提案されている．

- ***Pk*-匿名性**：あるテーブルについて，個人とレコードの対応関係を高々 $1/k$ の確率でしか推定できないとき，そのテーブルは Pk-匿名性を満たすという [15]．Pk-匿名性を満たす加工方法を Pk-匿名化とよぶ．Pk-匿名性は k-匿名性を確率的指標に拡張したものであり，PRAM やノイズ付加などによって加工されたテーブルについても識別リスクの評価が可能となる．表 4 は PRAM およびノイズ付加を用いて Pk-匿名化を行ったテーブルのイメージである．以下，PRAM の一手法である維持置換撹乱 (retention-replacement perturbation) [16] を例に Pk-匿名化を具体的に説明する．維持置換撹乱では，各属性 A_j について，属性値 $v \in A_j$ が $v' \in A_j$ に変化する確率を

$$p_{j,v,v'} = \begin{cases} \rho_j + \frac{1-\rho_j}{|A_j|} & (v = v') \\ \frac{1-\rho_j}{|A_j|} & (v \neq v') \end{cases}$$

と定義する．ρ_j は維持確率とよばれるパラメータである．たとえば $A_j = \{F, M\}$, $\rho_j = 0.5$ とすれば，F は 75%の確率で維持され，25%の確率で M に変化する．そしてレコード数を n としたとき，

表 3　2-匿名化されたテーブルの例

sex（準識別子）	age（準識別子）	income（センシティブ属性）	item 1（非センシティブ属性）
×	[24, 25]	$46K	coffee
×	[24, 25]	$52K	beer
M	30	$57K	cola
M	30	$81K	milk
F	32	$50K	cola
F	32	$104K	whiskey

表 4　*Pk*-匿名化されたテーブルのイメージ（太字は PRAM で変化した値，斜字はノイズ付加で変化した値）

sex（準識別子）	age（準識別子）	income（センシティブ属性）	item 1（非センシティブ属性）
F	24	$46K	coffee
F	*23*	$52K	beer
M	30	$57K	cola
M	*31*	$81K	milk
M	32	$50K	cola
F	*35*	$104K	whiskey

$$k \leq 1 + (n-1)\left(\prod_j \frac{1-\rho_j}{1+(|A_j|-1)\rho_j}\right)^2$$

となるよう ρ_j を選べば Pk-匿名性を満たすことが示されている [15]．ノイズ付加を用いた Pk-匿名化についても提案されている [17]．

3. 課題

2 節で挙げたような匿名化技法やリスク指標は，実際に利用する際いくつかの課題がある．特に問題となるのが，匿名化技法を適用したテーブルが所望の分析を行うに足るものかということであろう．たとえば global recoding や local recoding を用いて年齢を年代に置き換えた場合，5 歳刻みや未成年は細かい年齢で分析したい場合などに用いることが難しい．PRAM，ノイズ付加，スワッピングなどの撹乱的な手法は，元の値の細かさを維持できる利点があるが，分析の正確度を大きく損ねるかもしれない．この問題に対し，撹乱的な手法の影響を取り除きながら元データの統計量を推定する再構築法 (reconstruction method)，たとえば元データの生成分布を表す確率密度関数（確率分布）を推定する逐次ベイズ法 [16] が提案されている．さらには，確率密度関数を出力した際の正確度を理論的に保証する手法 [18] も提案されるなど，匿名化されたテーブルの有用性を高める研究が進展している．

その他の課題として，「次元の呪い」や「逐次開示」が挙げられる [19]．次元の呪いは，テーブルの属性数が増えるほどデータの一意性が高まるため，より強力な匿名化技法を用いる必要があり，結果，テーブルの有用性を著しく低下させてしまうという問題として知られる [20]（属性数の増加に伴う計算量の問題を指すこともある）．次元の呪いの基本的な対策は，テーブルを複数に分割して各々を匿名化することだろう．分析に応じて必要な属性のみを抽出する方法であり，たとえば表 1 の例では，{sex, age, income} や {age, income, item 1} など，重複を含めてテーブルを分割する．すべての属性を必要とするような分析においても，分析過程では一部の属性のみで十分な場合に有効となる．このように同一のパーソナルデータについて重複を含めて複数の匿名化されたテーブルを提供する場合は，提供したテーブル全体の依存関係を考慮した匿名性を評価する必要があり，これまでいくつかの指標が提案されている [21, 22]．

逐次開示の問題は，最新データと過去のデータを比較する場合などに生じる．同一個人のデータが動的変化するとき，変化前のデータと変化後のデータがそれぞれ匿名化されていても，変化前後に関する情報を知っていれば特定個人のデータの推定が可能となる場合がある [23]．逐次開示の対策についてはすでに多くの研究成果があるが [24, 25]，一般に有用性を損ねやすくなるため，有用性の向上が今後の課題といえる．

4. 属性開示リスクの評価指標

2 節で述べたように，識別が困難であっても，（非）センシティブデータの分布や背景知識によっては属性開示リスクが残る．具体的なリスクとして同種攻撃 (homogeneity attack) と背景知識攻撃 (background knowledge attack) が挙げられている [26]．同種攻撃は，匿名化されたデータのセンシティブデータの分布から特定の個人（ターゲット）のセンシティブデータを推定する攻撃であり，たとえばターゲットの準識別クラスのセンシティブデータがすべて同じであれば，識別はできなくてもターゲットのセンシティブデータを知り得てしまう．背景知識攻撃は，ターゲットの準識別クラスについて，センシティブデータに関する背景知識を利用してターゲットのセンシティブデータを推定する攻撃である．背景知識はたとえば，ターゲットの年収は「500 万円未満」などの事前に知っている情報である．

同種攻撃を考慮したリスク指標として，(α, k)-匿名性が提案されている [27]．任意の準識別クラスについて，任意のセンシティブデータの出現頻度の割合が α 以下かどうかを指標とする．(α, k)-匿名性を満たすための匿名化として，一般化が例示されている [27]．一方，背景知識攻撃を考慮したリスク指標として，p-センシティブ k-匿名性が提案されている [28]．任意の準識別クラスのセンシティブデータが p 種類以上存在するかどうかを指標とする．攻撃者がターゲットのセンシティブデータについて $p-2$ 種類までの可能性を排除できたとしても特定できない．背景知識攻撃を考慮した ℓ-多様性も提案されている [26]．具体的には 3 種類の指標を与えており，たとえば帰納的 (c, ℓ)-多様性は，c を定数，f_i をセンシティブ属性 A の i 番目に頻度の高いセンシティブデータの頻度としたとき，

$$f_1 < c \sum_{\ell \leq u} f_u$$

を満たすかどうかを指標とする．さらに帰納的 (c, ℓ)-多様性を確率的指標に拡張した帰納的 $P(c, \ell)$-多様性

も提案されている [29].

また，匿名化されたデータの開示前後における攻撃者の知識の差を確率的に評価する ρ_1-to-ρ_2 プライバシー侵害が提案されている [30]．具体的には，X をセンシティブデータの分布に関する確率変数，Y を匿名化後のセンシティブデータの分布に関する確率変数，$Q(x)$ をある述語関数としたとき，$0 < \rho_1 < \rho_2 < 1$ を満たす定数 ρ_1, ρ_2 について

$$\Pr(Q(X)) \leq \rho_1 \wedge \Pr(Q(X)|Y = y) \geq \rho_2$$

であれば（上向き）ρ_1-to-ρ_2 プライバシー侵害とよぶ．すなわち，$Q(X)$ が真となる確率が匿名化されたデータ y の開示によって ρ_1 以下から ρ_2 以上になるとリスクが高いとみなす．[30] では，ρ_2 以上から ρ_1 以下になる場合（下向き）も同様にリスクが高いとしている．

さらに攻撃例として歪み攻撃 (skewness attack) が挙げられ，当該攻撃を考慮した指標として t-近似性が提案されている [31]．歪み攻撃は，ターゲットの準識別クラスのセンシティブデータの分布とレコード全体のセンシティブデータの分布との差異を利用した攻撃であり，t-近似性は当該二つの分布の距離が一定以下であるかどうかを指標とする．ρ_1-to-ρ_2 プライバシー侵害に近い考え方といえる．

最後に，攻撃者の事前知識を仮定しないリスク指標として近年注目を集めている差分プライバシー (differential privacy) [32] を簡単に紹介する．差分プライバシーは，対話型データベースにおいて「特定個人のデータがデータベースに入っていてもいなくても開示される出力がほぼ変化しない」ことを指標とする．質的なデータからなるテーブルは（高次元の）度数表とみなすことができ，度数表において差分プライバシーを満たす手法が提案されている [32].

上記のように，これまでさまざまな属性開示リスクの評価指標が提案されている．その背景として，属性開示リスクに対する完全な保護は不可能であり [26, 32]，択一的な指標の確立が困難であること，そして識別リスクの対策以上に有用性を損ねやすいことが挙げられる．今後，適切で受容性の高い属性開示リスクの評価指標の確立が望まれる．

5. まとめ

古典的な匿名化では守れない個人特定やプライバシー侵害のリスクについて触れ，特に個人特定のリスクを低減させる匿名化技術とその課題について概説した．近年，プライバシーに配慮しつつパーソナルデータの二次利用を促進させる試みが国内外でみられるようになったが，セキュリティ対策やプライバシーの配慮はもちろんのこと，提供者である個人が安心できるような匿名化技術の確立が今後強く望まれる．同時に，パーソナルデータの保護一辺倒だけでなく，匿名性と有用性を高いレベルで両立できる技術の研究開発も今後ますます重要になるだろう．

参考文献

[1] L. Sweeney, “k-anonymity: A model for protecting privacy,” *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, **10**(5), pp. 557–570, 2002.

[2] L. Sweeney, “Uniqueness of simple demographics in the U.S. population,” Technical Report LIDAP-WP4, Carnegie Mellon University, Laboratory for International Data Privacy, 2000.

[3] 竹村彰通，“個票開示問題の研究の現状と課題，” 統計数理，**51**(2), pp. 241–260, 2003.

[4] D. Lambert, “Measure of disclosure risk and harm,” *Journal of Official Statistics*, **9**(2), pp. 313–331, 1993.

[5] 独立行政法人統計センター，「統計データ開示抑制に関する用語集 改訂版（対訳）」, http://www.nstac.go.jp/services/pdf/skk-yogosyu2.pdf（2016 年 3 月 25 日閲覧）

[6] 内閣官房，「技術検討ワーキンググループ報告書（第 5 回）パーソナルデータに関する検討会配布資料」, https://www.kantei.go.jp/jp/singi/it2/pd/dai5/siryou2-1.pdf（2016 年 3 月 25 日閲覧）

[7] P. Kooiman, L. C. R. J. Willenborg and J. M. Gouweleeuw, “PRAM: A method for disclosure limitation of microdata,” Report, Department of Statistical Methods, Statistics Netherlands, 1997.

[8] 藤野友和，垂水共之，“PRAM の理論とその実用上の諸問題，” 統計数理，**51**(2), pp. 321–335, 2003.

[9] L. Willenborg and T. de Waal, *Statistical Disclosure Control in Practice* (*Lecture Notes in Statistics* 111), Springer, 1996.

[10] L. Willenborg and T. de Waal, *Elements of Statistical Disclosure Control* (*Lecture Notes in Statistics* 155), Springer, 2001.

[11] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, R. Lenz, J. Naylor, E. S. Nordholt, G. Seri and P.-P. De Wolf, “Handbook on statistical disclosure control (version 1.2),” http://neon.vb.cbs.nl/casc/handbook.htm（2016 年 3 月 25 日閲覧）

[12] 佐井至道，“個票データにおける個体数とセル数との関係，” 応用統計学，**27**(3), pp. 127–145, 1998.

[13] A. Meyerson and R. Williams, “On the complexity of optimal k-anonymity,” In *Praceedings of PODS 2004*, pp. 223–228, 2004.

[14] C. Aggarwal and P. Yu, *Privacy-preserving Data mining: Models and Algorithms*, Springer, 2008.

[15] D. Ikarashi, R. Kikuchi, K. Chida and K. Takahashi, “k-anonymous microdata release via post randomisation method,” *Advances in Information and Computer Security* (*Lecture Notes in Computer Science* 9241), K. Tanaka and Y. Suga (eds.), Springer, pp. 225–241, 2015.

[16] R. Agrawal, R. Srikants and D. Thomas, “Privacy

preserving OLAP," In *Proceedings of SIGMOD 2005*, pp. 251–262, 2005.
[17] 五十嵐大，長谷川聡，納竜也，菊池亮，千田浩司，“数値属性に適用可能な，ランダム化により k-匿名性を保証するプライバシー保護クロス集計，”コンピュータセキュリティシンポジウム 2012 (CSS2012)，2012.
[18] 長谷川聡，正木彰伍，濱田浩気，菊池亮，“確率的 k-匿名化における再構築の正確度に関する理論的解析，”2016 年暗号と情報セキュリティシンポジウム (SCIS2016)，2016.
[19] 千田浩司，五十嵐大，高橋克巳，濱田浩気，菊池亮，冨士仁，“集合匿名化クラウドの課題と対策，”電子情報通信学会論文誌，**J96-A**(4), pp. 149–156, 2013.
[20] C. Aggarwal, "On k-anonymity and the curse of dimentionality," In *Proceedings of VLDB 2005*, pp. 901–909, 2005.
[21] D. Kifer and J. Gehrke, "Injecting utility into anonymized datasets," In *Proceedings of SIGMOD 2006*, pp. 217–228, 2006.
[22] 五十嵐大，千田浩司，濱田浩気，菊池亮，“秘匿計算とランダム化によるハイブリッド匿名化システム，”2012 年暗号と情報セキュリティシンポジウム (SCIS2012)，2012.
[23] K. Wang and B. Fung, "Anonymizing sequential releases," In *Proceedings of SIGKDD 2006*, pp. 414–423, 2006.
[24] B. Chen, D. Kifer, K. LeFervre and A. Machanavajjhala, "Privacy-preserving data publishing," *Foundations and Trends in Databases*, **2**(1–2), 2009.
[25] B. Fung, K. Wang, A. Fu and P. Yu, *Introduction to privacy-preserving data publishing*, CRC Press, 2011.
[26] A. Machanavajjhala, J. Gehrke, D. Kiefer and M. Venkitasubramanian, "ℓ-diversity: Privacy beyond k-anonymity," *ACM Transcations on Knowledge Discovery from Data*, **1**(1), Article No. 3, 2007.
[27] R. Wong, J. Li, A. Fu and K. Wang, "(α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing," In *Proceedings of ACM SIGKDD 2006*, pp. 754–759, 2006.
[28] T. M. Truta and B. Vinay, "Privacy protection: p-sensitive k-anonymity property," In *Proceedings of ICDE 2006*, 2006.
[29] 五十嵐大，千田浩司，高橋克巳，“$P\ell$-多様性—属性推定に対する再構築法のプライバシーの定量化—，”コンピュータセキュリティシンポジウム 2010 (CSS2010)，2010.
[30] A. Evfimievski, J. Gehrke and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," In *Proceedings of PODS '03*, pp. 211–222, 2003.
[31] N. Li and T. Li, "t-closeness: Privacy beyond k-anonymity and ℓ-diversity," In *Proceedings of ICDE 2007*, 2007.
[32] C. Dwork, "Differential privacy," *Automate, Language and Programming* (*Lecture Notes in Computer Science* 4052), M. Bugliesi et al. (eds.), Springer Berlin Heidelberg, pp. 1–12, 2006.

