情報処理学会論文誌 Vol.57 No.9 1900–1910 (Sep. 2016)

招待論文

匿名加工・再識別コンテスト Ice and Fire：匿名加工方式とその安全性を評価する試み

菊池 浩明[1,a)]

受付日 2016年2月8日, 採録日 2016年6月30日

概要：個人情報の保護とビッグデータの活用を両立させるために，個人情報の匿名加工の法整備が進み，様々な方式が導入されようとしている．その一方，匿名加工の為の方式には様々な提案が行われており，適用する分野やデータの種別に適した方式をどのように選定するか定まっていない．加えて，有用性と安全性を正しく評価する方法が確立していない．そこで，共通の疑似データを用いて，匿名加工とその再識別の技術を競うコンテストを企画する．本稿では，このコンテストの目的，提供する疑似データの選定，サンプルの匿名加工と再識別アルゴリズム，有用性評価方法，安全性評価方法について述べる．

キーワード：個人情報，匿名加工，仮名，再識別，k-匿名性

Ice and Fire: Data Anonymization and De-Anonymization Competition

Hiroaki Kikuchi[1,a)]

Received: February 8, 2016, Accepted: June 30, 2016

Abstract: Data anonymization is ready to go before the big-data business runs successfully while preserving privacy of personal information. While, it is not trivial to choose the best algorithm to make the given data anonymized to be secure for a given particular purpose. To access the risk to be compromised accurately, the data needs to balance the utility and the security. Hence, with common pseudo micro-data, we propose a competition for best anonymization and re-identification algorithm. The paper addresses the aim of the competition, the target micro-data, sample algorithms, utility and security metrics.

Keywords: personal information, de-identification, pseudonym, re-identification, k-anonymity

1. はじめに

スマートフォンの普及とネットワーク技術の発達は，私たちの生活を大きく変えた．常時オンラインとなった端末からのリアルタイムなメッセージの交換，蓄積された膨大なデータからのトレンド分析，位置情報や個人の嗜好を反映した質の高いサービス，行動を推測した効率の良い管理．これらのビッグデータを活用した新しい IT 技術は，その便利さの一方で，個人のプライバシーを損なう危険性を生じさせている．たとえば，スマートフォンからの位置情報を取り続けることでその人の自宅の住所や交友先の関係先の情報を調べたり，カメラで取得した顔画像からその人の年齢や性別を予測したり，顔画像データベースと照合してその人が誰かを特定したりすることも可能になってきた．さらに，これらの膨大なデータを利活用するためには，異業種間でのデータの交換が必要にもかかわらず，個人情報保護の枠組みではデータの第三者提供に関する取り決めが十分ではなかった．

このような社会的な要請を受けて，2015 年 9 月，我が国の個人情報保護法[*1]が改正された．2016 年 1 月には個人情

[1] 明治大学総合数理学部
School of Interdisciplinary Mathematical Sciences, Meiji University, Nakano, Tokyo 164–8525, Japan
[a)] kikn@meiji.ac.jp

本論文は，文献 [1] 菊池他，“匿名加工・再識別コンテスト Ice & Fire の設計” を基にしている．

[*1] 「個人情報の保護に関する法律及び行政手続きにおける特定の個人を識別するための番号の利用等に関する法律の一部を改正する法律（平成 27 年 9 月 9 日法律第 65 号）」平成 27 年 9 月 3 日成立，同月 9 日公布

表 1 改正個人情報保護法
Table 1 Amended Act on the Protection of Personal Information.

(1) 個人情報の定義の明確化. 個人識別符号（第 2 条 2 項），要配慮個人情報（第 2 条 3 項）
(2) 個人情報の有用性確保. 匿名加工情報（第 2 条 9 項），認定個人情報保護団体による個人情報保護指針（第 53 条）
(3) 個人情報の保護を強化. 個人情報データベース等提供罪（第 83 条）
(4) 個人情報保護委員会の新設と権限（第 5 章第 59 条，74 条）
(5) 個人情報の取り扱いのグローバル化 国境を越えた適用（第 75 条），外国執行局への情報提供（第 24 条）
(6) オプトアウトの扱い. 届出厳格化（第 23 条 2 項），利用目的の変更禁止（第 15 条 2 項）

報保護委員会が発足し，2017 年には各省庁で管理指導している保護を一元化する予定である．**表 1** に改正法の主な項目を列挙する．オプトアウトの厳格化，グローバル化への対応に並び，大きな柱となっているのが，「匿名加工情報」の新設である．「(定められた措置を講じて）特定の個人を識別することができないように個人情報を加工して得られる個人に関する情報であって，当該個人情報を復元することができないようにしたもの（第二条 9 項)」と定められている．匿名加工情報は，あらかじめ情報の項目と提供方法を記録し，トレーサビリティを保証することで，本人の同意がなくても第三者に提供することが許されており，IoT をはじめとするビッグデータの活用を推進する概念として期待されている．

しかしながら，その加工方法については標準的な定めはなく，業種ごとに定められた 44 の認定個人情報保護団体が，消費者や関係者の意見を聴いて，その対象事業に適した個人情報保護指針を定めること（第五十三条）とされている．各業界のデータの種類や特性は多様で，漏えいのリスクや被害の度合いもまちまちなため，業界に特有の安全管理措置と匿名加工方法を定めることとしたものと考えられる．しかし，たとえ業種や用途を特定したとしてもデータの加工方法は自明ではない．たとえば，[9] では，経済産業省が野村総合研究所に委託して 2015 年 3 月に発表した「パーソナルデータ利活用に関するマルチステークホルダープロセスの実施方法等の調査事業」で検討した匿名加工における課題が報告されている．クレジットカード利用者の購買データを匿名加工して，新規顧客を開拓しようとする 7 つの事例のうち，認められたのは条件付きの 1 つだけであった．

世界的な個人情報保護の動きも加速している．欧州委員会は，2016 年 4 月に一般データ保護規則 General Data Protection Regulation（GDPR）を可決した．2012 年の初案の提示から 4 年間を経てようやく可決に至ったもので，今後規則内容を具体化して 2018 年に施行の予定とされている．ウェブの検索結果の異議申し立てを認める「忘れられる権利」や，個人のプロファイリングを規制する権利など，新たな権利が明文化された一方で，匿名加工のための基準や実質的なルール形成については，各国のプライバシーコミッショナーに委ねられている．法的なフレームワークと実装可能なプライバシー保護技術や評価指標との間にギャップがあることを指摘し，その溝を埋めるための行為規範が提案されている [7].

ISO/IEC の標準化合同委員会 JTC 1/SC27 の WG 5（Privacy, Identity management and biometrics）においては，29100 Privacy framework, 29101 Requirements for partially anonymous, partially unlinkable authentication, 29101 Privacy architecture framework, 27018 Code of practice for PII protection in public clouds acting as PII processors などの標準化が行われた．なかでも，現在策定中の 20899（Privacy enhancing data de-identification techniques）においては，*specifing infroamtion*（特定の個人を識別する情報）と *singling out but not specifying information*（識別はできるが特定されていない情報）の 2 つを明確に区別することが検討されている [10]．後者は，パーソナルデータ検討会技術検討 WG で提案されていた「識別非特定情報」に該当するものであり，本稿で議論する匿名加工に密接に関係している．これは大きな前進であるが，de-specifying（非特定化）や de-singling out（非識別化）の具体的手法や基準についてはまだ藪の中である．

匿名加工の方法が明確に定まっていないことの原因として，匿名加工方法の安全性と有用性を評価するためのテストベッドとなる共通データの欠乏があると考えている．研究目的として人工的に機械生成したデータには，現実の問題では避けられない極端な偏りや欠損などが排除されていたり，値に意味がないために漏えいの危険性がイメージしにくい．再識別のリスクを評価しようにも，しばしばそのような試みは危険な行為とみなされて敬遠されているところがある．

そこで，情報処理学会コンピュータセキュリティ研究会では，プライバシーワークショップ実行委員会を組織し，匿名加工と再識別のコンテスト “Ice & Fire” [*2] を 2015 年 10 月 21 日，長崎ブリックホールにおいて開催した．本コンテストの概要を**表 2** に示す．本コンテストの目的は，

- 高い安全性と有用性を有した匿名加工技術の開発
- 再識別に対する公平な安全性評価手法の確立

にある．匿名加工の対象は，教育機関などの演習用として独立行政法人統計センターが作成した疑似ミクロデータ [8] である．本データは，平成 16 年全国消費実態調査を基に

[*2] Ice は匿名加工処理，Fire は再識別処理を暗示し，両者がそれぞれ技巧を競うことを想定している．

表 2 匿名加工コンテスト概要
Table 2 Outline of the Anonimizing Technology Competition.

予備戦（匿名加工フェーズ）	8 月 24 日–9 月 24 日
予備戦（再識別フェーズ）	9 月 25 日–10 月 9 日
本戦	10 月 21 日 10:00–12:15
最終プレゼンテーション	10 月 22 日
場所	長崎ブリックホール
主催	プライバシーワークショップ実行委員会
参加者	予備戦 17 チーム（81 名），本戦 12 チーム

して生成されており，現実の統計値とほぼ同等の性質を有し，任意の目的で自由に加工することが許されている．したがって，改正法や利用環境に制限されることなく，あらゆる手法で再識別を試みることができる．こうして，様々な匿名加工技術を共通の環境で，公平に定量評価することにより，匿名加工の課題を明らかにし，より高い有用性と安全性を持つ匿名加工方式の開発を促進する．

本稿は，本コンテストで用いる技術の基本定義と有用性と安全性の評価指標を述べる．コンテストの設計においては，考慮した不正行為と対象としたリスクについて述べ，サンプルとするいくつかの再識別アルゴリズムを定義する．

2. 匿名加工アルゴリズムと再識別

2.1 匿名加工とリスク

匿名化（anonymization）とは，Emam は ISO/TS 25237 を引いて「データ主体とそれを識別するデータの相関を取り除く処理」と定めている [3]．再識別（re-identification）は，匿名加工データを分析したり，他のデータセットと照合することにより，匿名加工データ[*3]と元の個人データを結びつける処理を指す．ただし，ここには様々なリスクを含むあいまいさがあり，それらの脅威の例を，**図 1** に示す．ここで，氏名，年齢，購買履歴を含むデータ（2 つのレコードの表）が個人情報であり，名前をランダムな “320” などの仮名にしたり，年齢を年代に一般化したりして作成したデータが匿名加工情報を表す．匿名加工に関して，次のような脅威が存在する．

1. **特定** 匿名加工されたデータが復元され，特定の個人を同定すること．
2. **識別** 複数のデータが，ある（特定ではない）個人の情報であることを識別すること．レコードのリンクや履歴からの追跡が含まれる．
3. **属性推定** 加工されたデータから，本人の一部の属性を推定すること．
4. **本人へのアクセス** ダイレクトメールなどの広告などが本人に届くこと．
5. **他のデータとの照合** レコードの突合せを目的として，独立した他のデータと匿名加工データを照合すること．

[*3] 個人情報を体系的に構成し，データベースなどに格納したものを個人情報データベース，データベースを成す値を個人データと呼ぶことに習い，本稿でも，体系的に構成された匿名加工情報であることを暗黙的に示すように，匿名加工データと呼ぶ．

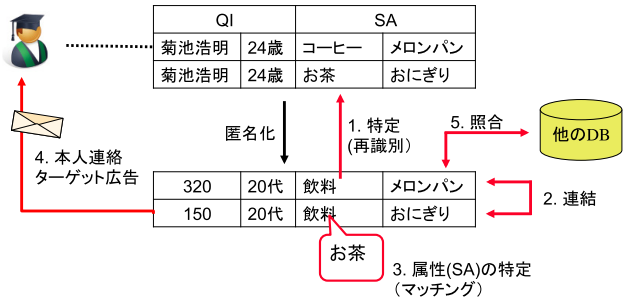

図 1 匿名加工と再識別リスク
Fig. 1 Risks of anonymized data.

改正法で論じられているのは，1 の特定と 5 の照合のみである．1 は，第二条 9 項の匿名加工情報の定義「特定の個人を識別出来ない」と「当該個人情報を復元することが出来ない」との条件で定められており，5 は第三十八条の「（本人を識別するために）他の情報と照合してはならない」で禁じられている．しかし，2 や 3 に関しては該当する項目がなく，むしろ，第二条 9 項の括弧書き（規則性を有しない方法により他の記述などに置き換えることを含む）が暗示するように，識別されていても匿名加工情報とみなすように解釈できる．

そこで，本コンテストでは，個人の再識別リスクに焦点を置き，後述するようにレコードのリンクで再識別の度合いを定量化した．

2.2 基本定義

データセット $\boldsymbol{X}$ は，m 個の属性 $X^1,\ldots,X^m$ を持つ n 個のレコード $\boldsymbol{x}_i = (x_i^1,\ldots,x_i^m)$ $(i = 1,\ldots,n)$ からなる．レコード行番号 $I^X = (1,\ldots,n)$ は，各レコードの行番号表す．属性は，連続値，順序属性値，カテゴリ属性値，ブール値などがあり，属性 X の値域を $R(X)$ で表す．データセット $\boldsymbol{X}$ は，特定の個人を識別する（特定する）情報を含むとき，その時に限り個人情報データセットとなる．

本稿では，データセット $\boldsymbol{X}$ から処理された匿名加工データを $\boldsymbol{Y}$ で表す．匿名加工データ $\boldsymbol{Y}$ は，$\{X^1,\ldots,X^m\}$ の部分集合の $m'(m' \le m)$ 個の属性についての $n'(n' \le n)$ 個のレコード $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{n'}$ から成る．$\boldsymbol{Y}$ のレコード $\boldsymbol{y}_j = (y_j^1,\ldots,y_j^{m'})$ は，データセット $\boldsymbol{X}$ のあるレコード $\boldsymbol{x}_i$ を基にして処理されており，その関係を $j = \pi(i)$ となる関数

$$\pi : \{1,\ldots,n\} \to \{1,\ldots,n'\}$$

で表す．π は全射（トップコーディングなどにより識別されやすいレコードを排除することがあるため）や単射（1 つのレコードから複数の匿名加工レコードが生成されることとあるため）とは限らないので，置換（permutation）となる保証はないことに注意が必要である．匿名

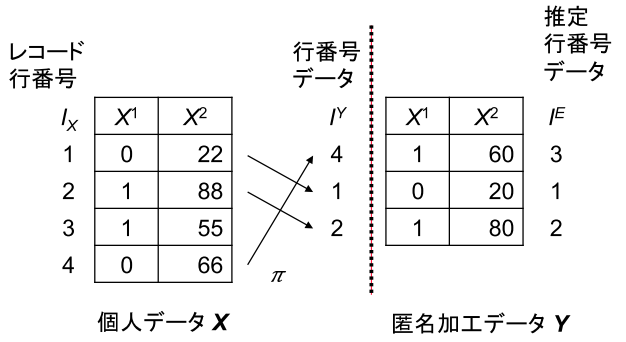

図 2 行番号データ
Fig. 2 Record index sequences.

加工データにおけるレコードの行番号を関数 π を用いて $I^Y = (i_1^Y, \ldots, i_{n'}^Y) = (\pi^{-1}(1), \ldots, \pi^{-1}(n'))$ で表す（図 2 参照）．

$\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ をデータセット $\boldsymbol{X}$ に属する 2 つのレコードとする．すべての $X^q \in QI$ について，$x_i^q = x_j^q$ である時，$\boldsymbol{x}_i$ と $\boldsymbol{x}_j$ は QI-同値関係にあると呼び，$\boldsymbol{x}_i \approx^{QI} \boldsymbol{x}_j$ と表す．QI 同値関係について，データセット $\boldsymbol{X}$ は $R(QI)$ の要素数 $|R(QI)|$ 個の同値類に直和に分割される．ただし，存在しない QI の組は除く．

本稿では，狭義の再識別として，匿名加工データ $\boldsymbol{Y}$ から，$\boldsymbol{X}$ のレコード行番号を推定することと定める．すなわち，攻撃アルゴリズム E により，行番号データ I_Y と等しくなる（再識別）推定行番号データ $I^E = (i_1^E, \ldots, i_{n'}^E) \cong I^Y$ を求めることである．以上の行番号データと推定行番号データの関係を図 2 に示す．この例では，$n = 4$ の個人データから，行削除とレコード置換が行われ，$n' = 3$ の匿名加工データが生成されている．推定行番号データ I^E は 1 行目 (3) が誤っている．

2.3 準識別子 QI と機微属性 SA

属性 X には，マイナンバーのような直接的な識別子，性別，住所，年齢などのような，組み合わせることで個人を識別することが可能な間接的な識別子である準識別子（quasi-idenitifier，以後 QI），そして，それ以外の属性がある．それ以外の属性の内，病名や思想信条などの配慮な必要な情報を機微属性（sensitive information，以後 SA）[*4] と呼ぶ．Emam は，QI と SA の例として，それぞれ，学歴，日常言語，学歴，離婚歴などと，遺伝的情報，精神疾患歴，性的情報，生殖情報などを上げている（文献 [3] 2 章のコラム）．

しかしながら，機微の度合いには個人差が大きく，QI とも SA とも考えられる属性もあり，その定義はしばしば議論になる．そこで本稿では，属性 $X^1, \ldots, X^m$ の内，静的な情報で，しばしばカテゴリー化されている離散値を取る属性を QI と見なし，それ以外をすべて SA と考える．QI と SA の属性を集合 $QI, SA \subset \{1, \ldots, m\}$ で表す．$QI \cup SA = \{1, \ldots, m\}$, $QI \cap SA = \emptyset$ である．

[*4] コンテストで QI 以外の属性につけた名前であり，改正法でいう「要配慮情報」ではない．

2.4 攻撃者

再識別を行う攻撃者に許されるリソースは，様々な場合が考えられ，一様な仮定は困難である．たとえば，Emam は，攻撃の種類として次の 4 つの場合を上げている [3]．

- T1：故意の試み（deliberate attempt），
- T2：不慮の試み（inadvertent attempt），
- T3：規則違反（data breach），
- T4：公開情報（public data）．

偶然に対象とする個人の情報を入手する確率 Pr（acquintance）は低いが，入手した時に匿名加工からの再識別できる条件付き確率 Pr（re-id | acquaintance）は高く，T2 が成立する同時確率はそれらの積で決まる．

英国プライバシーコミッショナー ICO は，T1 の仮定を提言している [2]．「動機付けられた攻撃者」は，いかなる予備知識も持たないが，匿名加工データからある特定の個人のレコードを特定しようと企てる攻撃者である．政治家や著名人に対してウェブ検索やソーシャルネットワーキングサービスなどのあらゆる公開情報を用いる例が該当する．

Domingo は，背景知識を仮定することが困難なことに対して，オリジナルデータ $\boldsymbol{X}$ と匿名加工データ $\boldsymbol{Y}$ の両方を持ち，$\boldsymbol{X}$ から $\boldsymbol{Y}$ の写像を復元することを試みる最大知識攻撃者モデル（maximum-knowledge attacker）を導入して，データ漏えいのリスクを評価している [6]．

一方，我が国の匿名加工情報は，いわゆる「提供元基準」であり，攻撃者の能力や背景知識に依存せず，いかなる第三者に提供し流通したとしても，再び個人を識別しないように加工することが求められている．したがって，本コンテストの攻撃者は最大の知識，すなわち，オリジナルの個人データ $\boldsymbol{X}$ と照合して，匿名加工データ$\boldsymbol{Y}$の再識別を試みることとした．

2.5 匿名加工アルゴリズム例

匿名加工の技術として，

(1) 属性（列）削除
(2) レコード（行）削除
(3) セル（値）削除
(4) 一般化
(5) ミクロアグリゲーション（micro-aggregation）
(6) データスワッピング（data swapping）
(7) ランダマイズドレスポンス（randomized response）
(8) 加法摂動化（additive perturbation）
(9) リサンプリング（resampling）

がよく知られている．これらの技術を組み合わせて，k-匿名性，ℓ-多様性，t-近似性などの安全性指標を満たすよう

に個人データを匿名加工する．

2.6 再識別アルゴリズム例

再識別アルゴリズムには標準的なものは知られていない．攻撃者に与えられる知識の想定ができないため，共通のアルゴリズムを考えにくいためである．そこで，本コンテストでは，オリジナルの個人データ $\boldsymbol{X}$ を与えて，匿名加工データ $\boldsymbol{Y}$ から行番号データ I^Y を推定する次のアルゴリズム [1] を用意した．

Sort 小さな摂動化に対しては強力な再識別を実行するが，ソートによって識別をしているためトップコーディングなどのレコード削除に対して弱い．

IdRand k-匿名性が低いデータセットをより高い確率で再識別する．一様な確率で推定するために，候補レコード集合 $C(\boldsymbol{y})$ が小さいほど識別率が上がる．安全性指標 k-anony の平均値が $\overline{k}$ である時，この攻撃アルゴリズムが再識別するレコード数の期待値は $2/\overline{k}$ である．

IdSA IdRand と同様に，候補レコード集合 $C(\boldsymbol{y})$ を用意するが，指定された SA のある属性について $\boldsymbol{y}$ に最も近いレコードを選ぶところが異なる．

Algorithm 1 再識別アルゴリズム Sort [1]

(1) 入力：匿名加工データ $\boldsymbol{Y}$，個人データ $\boldsymbol{X}$，
出力：推定行番号データ $I^{Sort} = (i_1^{Sort}, \ldots, i_{n'}^{Sort})$

(2) レコード $\boldsymbol{x}_j$ の特徴量を，そのレコードの SA の和で $s_j = \sum_{i \in SA} x_j^i$ と定める．

(3) 個人データ $\boldsymbol{X}$ と匿名加工データ $\boldsymbol{Y}$ をそれぞれ，上記の特徴量についてソートした時の順位を $Rank^X$，$Rank^Y$ とする．

(4) $\boldsymbol{y} \in \boldsymbol{Y}$ の各レコードについて，

$$Rank^X(\boldsymbol{x}_k) = Rank^Y(\boldsymbol{y}_j)$$

となる順位 k を推定行番号データ $i_j^{Sort} = k$ とする．

Algorithm 2 再識別アルゴリズム IdRand [1]

(1) 入力：匿名加工データ $\boldsymbol{Y}$，個人データ $\boldsymbol{X}$，
出力：推定行番号データ $I^{IR} = (i_1^{IR}, \ldots, i_{n'}^{IR})$

(2) レコード $\boldsymbol{y}_i \in \boldsymbol{Y}$ について，QI を共通にする $\boldsymbol{X}$ の候補レコード集合 $C(\boldsymbol{y}_i) = \{\boldsymbol{x} \in \boldsymbol{X} | \boldsymbol{x} \approx^{QI} \boldsymbol{y}_i\}$ を求める．

(3) $C(\boldsymbol{y})$ から一様な確率 $p = 1/|C(\boldsymbol{y})|$ でレコード $\boldsymbol{x}_{j*}$ を選び，推定行番号データ $i_i^{IR} = j*$ とする．すべての $\boldsymbol{y}_i$ について Step 2 から繰り返す．

3. コンテスト

3.1 疑似ミクロデータ

「(教育用) 疑似ミクロデータ（以後，GMD と呼ぶ)」は，公的統計のミクロデータの利用を図るため，教育機関などの演習用として独立行政法人統計センターが作成した疑似的なミクロデータである [8]．約 59,400 世帯（単身 5,002 世帯を含む）に対して 3 カ月間の家計簿を調べた平成 16 年全国消費実態調査を基に生成されている．この個票データから高次元の集計表を作成し，各セルの量的属性値が多変量（対数）正規分布に従うことを仮定して，多変量正規乱数を作成することで作成されており，元の個票データの特性を保存している．

Algorithm 3 再識別アルゴリズム IdSA [1]

(1) 入力：匿名加工データ $\boldsymbol{Y}$，個人データ $\boldsymbol{X}$，識別用属性 $X^s \in SA$
出力：推定行番号データ $I^{IS} = (i_1^{IS}, \ldots, i_{n'}^{IS})$

(2) レコード $\boldsymbol{y}_i \in \boldsymbol{Y}$ について，QI を共通にする $\boldsymbol{X}$ のレコードの集合 $C(\boldsymbol{y}_i) = \{\boldsymbol{x} \in \boldsymbol{X} | \boldsymbol{x} \approx^{QI} \boldsymbol{y}_i\}$ を求める．

(3) $C(\boldsymbol{y})$ の各レコード $\boldsymbol{x}_j$ について，属性 X^s について $\boldsymbol{y}_i$ との距離最小となるレコード $\boldsymbol{x}_{j*}$ を

$$j^* = \underset{j \in C(\boldsymbol{y}_i)}{\operatorname{argmin}} |x_j^s - y_i^s|$$

により定め，推定行番号データ $i_i^{IS} = j*$ とする．

Algorithm 4 再識別アルゴリズム SA21 [1]

(1) 入力：匿名加工データ $\boldsymbol{Y}$，個人データ $\boldsymbol{X}$，出力：推定行番号データ $I^{S_{21}} = (i_1^{S_{21}}, \ldots, i_{n'}^{S_{21}})$

(2) レコード $\boldsymbol{y}_i \in \boldsymbol{Y}$ の特徴量を $y_j^{S_{21}}$ とする．

(3) 個人データ $\boldsymbol{X}$ と匿名加工データ $\boldsymbol{Y}$ をそれぞれ，特徴量 $y_j^{S_{21}}$ についてソートした時の順位を $Rank^X$，$Rank^Y$ とする．

(4) $\boldsymbol{y} \in \boldsymbol{Y}$ の各レコードについて，

$$Rank^X(\boldsymbol{x}_k) - 1 = \lfloor (Rank^Y(\boldsymbol{y}_j) - 1) \times \frac{n}{n'} \rfloor$$

となる順位 k を推定行番号データ $i_j^{S_{21}} = k$ とする．

表 3 NSTAC 疑似ミクロデータ仕様 [8]
Table 3 Statistics for the NSTAC pseudo microdata.

データセット	レコード数	QI 数	SA 数	
			支出項目	収入項目
	n	m		
大規模データ	32,027	14	149	34
簡易データ	8,333	14	11	N/A

表 3 に，GMD の基本仕様を示す．簡易データは，大規模データの中から世帯人員が 4 名で有業人員が 1 名から 2 名の世帯のみの $n = 8{,}333$ レコードからなっている．属性の内，世帯主の年齢，住居の種類などの質的属性 14 項目を QI，食糧，住居，光熱，教育などに分類された消費支出と年間収入などの量的属性を SA とする．簡易データは，消費支出の十大費目のみを有する．

3.2 プレイヤー

本コンテストには次のプレイヤーがある．

(1) 匿名加工者（ディフェンス）GMD $\boldsymbol{X}$ を基に，匿名加工処理を行い，匿名加工データ $\boldsymbol{Y}$ と行番号データ I^Y を生成する．再識別者に $\boldsymbol{Y}$ を，審判員に $\boldsymbol{Y}$ と I^Y を提出する．

(2) 再識別者（オフェンス）匿名加工データ $\boldsymbol{Y}$ を受け取り，GMD $\boldsymbol{X}$ を参照して，再識別アルゴリズムを実

施して推定した推定行番号データ I^E を審判員に提出する．

(3) 審判員（ジャッジ）$\boldsymbol{Y}$ の有用性指標を評価する．I^Y を参照して，$\boldsymbol{Y}$ の安全性指標を評価する．I^E と I^Y を照合して，再識別率を算出する．

3.3 有用性指標の定義

3.3.1 有用性指標 meanMAE

$SA = \{14, 15, \ldots, 25\}$ についてのデータセット平均値を求め，オリジナル個人データ$\boldsymbol{X}$と匿名加工データ $\boldsymbol{Y}$とのMAE（平均絶対誤差）で定める．匿名加工データ $\boldsymbol{Y}$について，

$$\mathrm{meanMAE}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{m} \sum_{i \in SA} |\mu(X^i) - \mu(Y^i)|$$

とする，ただしここで，

$$\mu(X^i) = \frac{1}{n} \sum_{j=1}^{n} x_j^i,\ \mu(Y^i) = \frac{1}{n'} \sum_{j=1}^{n'} y_j^i$$

とする．

meanMAE は小さいほど有用性が高い．

3.3.2 有用性指標 cross

属性の部分集合 $A = \{a_1, a_2, \ldots\} \subset QI$ について，可能な値のすべての組合せについて，$B \in SA$ の値の平均（Mean）と集計数（Cnt）を求め，オリジナルの個人データ$\boldsymbol{X}$と匿名加工データ$\boldsymbol{Y}$の間の MAE で定める．すなわち，

$$\mathrm{crossMean}^{A,B}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{|R(A)|} \sum_{a \in R(A)} |\mu(\boldsymbol{x}^B | \boldsymbol{x}^A = a) - \mu(\boldsymbol{y}^B | \boldsymbol{y}^A = a)|$$

$$\mathrm{crossCnt}^{A,B}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{|R(A)|} \sum_{a \in R(A)} \big| |\{\boldsymbol{x} \in X | \boldsymbol{x}^A = a\}| - |\{\boldsymbol{y} \in Y | \boldsymbol{y}^A = a\}| \big|$$

とする[*5]．ここで，$R(A)$ は属性 $X^{a_1}, X^{a_2}, \ldots$ における値域の直積の部分集合であり，$Y^A = a$ を満たすレコードが $\boldsymbol{Y}$ に存在しない時，$\mu(\boldsymbol{y}^B | \boldsymbol{y}^A = a) = 0$, $|\emptyset| = 0$ とする．たとえば，$A = \{7, 9\}$（性別，就業），$B = 15$（消費支出）とすると，$R(A) \subset \{(1,1), (1,2), (1,V), (2,1), (2,2), (2,V)\}$ である．$a = (1,2)$ とすると，$\mu(X^B | X^A = a)$ は，$x^7 = 1$（男），$x^9 = 2$（就業）となるレコードだけに制限した X^{15} 消費支出の平均値である．

crossMean, crossCnt とも小さいほど有用性が高い．

この指標は A と B によって評価が変わる．年齢と就業の有無に応じて (A)，住居費 (B) がどれ位変動するかなどの評価をモデルしている．

[*5] 外側の | が絶対値，内側の | が集合の要素数であることに注意する．

表 4 $QI = \{7, 9\}$ の時の k-匿名性
Table 4 QI-equivalence classes for $QI = \{7, 9\}$.

性別 \ 就業	1	2	V
1	8,051	27	9
2	127	101	18

3.3.3 有用性指標：corMAE

SA の属性の任意の 2 組 X^i, X^j について，ピアソンの相関係数 cor を求め，個人データ $\boldsymbol{X}$ と匿名加工データ $\boldsymbol{Y}$ との間の MAE で定める．

$$\mathrm{corMAE}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{|SA|^2} \sum_{i,j \in SA} \big|\mathrm{cor}(X^i, X^j) - \mathrm{cor}(Y^i, Y^j)\big|$$

corMAE は小さいほど有用性が高い．

3.3.4 有用性指標 nrow

匿名加工アルゴリズムの中には，レコード削除のように，識別の容易な特異なレコードを削除するものがある．レコードを削除すればするほど再識別のリスクが下がる．しかし，そのことによる有用性の損失を，匿名加工データ $\boldsymbol{Y}$ のレコード数で，

$$\mathrm{nrow}(\boldsymbol{Y}) = |n - |\boldsymbol{Y}|| = |n - n'|$$

と定める．

nrow は小さいほど有用性が高い．

3.4 安全性指標の定義

3.4.1 安全性指標 k-anony

QI 同値類についての安全性を，

$$k\text{-anony}(\boldsymbol{X}) = \min_{a \in R(QI)} |\{\boldsymbol{x} \in \boldsymbol{X} | X^{QI} = a\}|$$

$$k\text{-anonyMean}(\boldsymbol{X}) = \frac{1}{|R(QI)|} \sum_{a \in R(QI)} |\{\boldsymbol{x} \in \boldsymbol{X} | X^{QI} = a\}|$$

とする．ここで，$X^{QI} = a$ は，QI のすべての属性 $X^{q_1}, X^{q_2}, \ldots$ の直積が a であるようなレコードを表す．すなわち，k-anony は，k-匿名性の k を，k-anonyMean はその平均値を与えている．

たとえば，$QI = \{7, 9\}$（性別，就業）について GMD のレコード $\boldsymbol{X}$ は，**表 4** に示される個数だけあったとする．値域はそれぞれ，$\{1, 2\}, \{1, 2, V\}$ である．このとき，k-anony $= 9$（$k = 9$），k-anonyMean $= 1{,}388.8$ である．

k-anony，Mean は大きいほど安全性が高い．

3.4.2 安全性指標 re-id

匿名加工データ $\boldsymbol{Y}$ とその行番号データ $I^Y = (i_1^Y, \ldots, i_{n'}^Y)$ とする．再識別アルゴリズム E によって，$\boldsymbol{Y}$ を分析して推定した推定行番号データを $I^E = (i_1^E, \ldots, i_{n'}^E)$ とする．この時，R による再識別率を，正しく識別できたレコードの

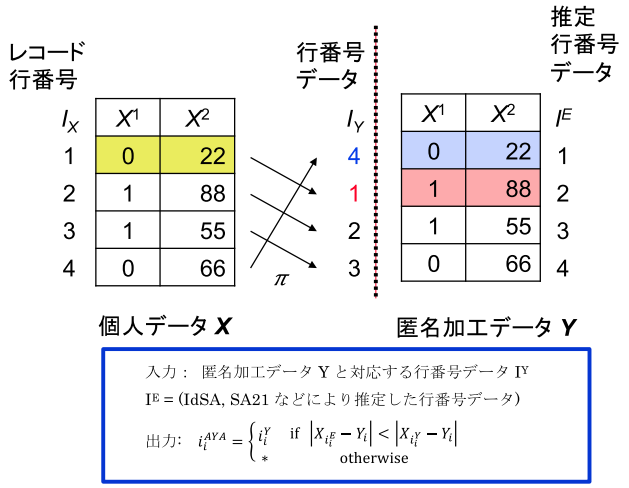

図 3 山岡匿名加工用再識別アルゴリズム AYA
Fig. 3 Re-identifying algorithm for cheating permutation AYA.

割合，すなわち，

$$\text{re-id}^E(I^Y, I^E) = \frac{\left|\{j \in \{1,\ldots,n'\} | i_j^Y = i_j^E\}\right|}{n'}$$

と定める．$\text{re-id}(I^Y, I^Y) = 1.0$ である．

再識別アルゴリズム Sort, IdRand, IdSA の再識別率をそれぞれ，re-id^{Sort}, re-id^{IdRand}, re-id^{IdSA} とする．

3.5 再識別率の課題「山岡匿名加工」

レコードの順序を変える匿名加工方法を文献 [1] の提案者の名前から「山岡匿名加工」と呼ぶ．たとえば，

$$\boldsymbol{y}_i = \begin{cases} \boldsymbol{x}_n & (i = 1) \\ \boldsymbol{x}_{i-1} & (i > 1) \end{cases}$$
$$I^Y = (1, 2, \ldots, n)$$

とするのが「山岡匿名加工」の例である．すなわち，「山岡匿名加工」とは，属性値は操作せずに行番号データだけを置換する処理である．なぜ単なる置換の処理を匿名加工の 1 つとして特別扱いするかというと，レコード識別を競う本コンテストにおいてのみ圧倒的な安全性と有用性を達成するからである．この匿名加工データは，普通の再識別アルゴリズムからは推定行番号データは $I^E = (n, 1, 2, \ldots, n-1)$ とされ，再識別率は 0 になる．有用性は，レコード順序を変えただけなので低下していない．実際にはデータに何も加工を加えていないにもかかわらず，安全性と有用性の両方で高く評価されてしまう．

そこで，本戦では，**図 3** で示す山岡匿名加工の匿名加工データを検出して，安全性評価値を下げる再識別アルゴリズムを導入した．ここで，匿名加工データ $\boldsymbol{Y} = \boldsymbol{X}$ であるにもかかわらず，行番号データ I_Y は偽りのデータである．対山岡匿名加工アルゴリズム AYA は，独自の方法（本戦では Sort が用いられた）で $\boldsymbol{Y}$ の再識別を行い，たとえば図の 1 行目 $(0, 22)$ ならば，最も近いと予測したレコード $X_1 = (0, 22)$ と匿名加工者が提出した行番号データ $I_Y(1) = 4$ 番目のレコード $X_4 = (0, 66)$ を比較し，推定したレコードとのユークリッド距離が小さければ，そのレコードは再識別されたと見なすというものである．

参加者にはこの AYA がサンプル再識別アルゴリズムとして加わることを事前に伝え，山岡匿名加工を抑制した．

3.6 有用性指標 IL

Domingo-Ferrer らは，文献 [5] にて匿名加工データの安全性を測る指標について提案している．この指標を基にして，個人データ $\boldsymbol{X}$ についての匿名加工データ $\boldsymbol{Y}$ とその行番号データ I^Y について，

$$IL(\boldsymbol{X}, \boldsymbol{Y}, I^Y) = \frac{1}{m'n'} \sum_{i=1}^{m'} \sum_{j=1}^{n'} \frac{|x^i_{i^Y_j} - y^i_j|}{\max x^i - \min x^i}$$

と定める．Domingo らの指標は，匿名加工によってレコードが削除することは仮定されていないので，変形して用いる．

IL は小さいほど有用性が高い．

3.7 総合評価

匿名加工データの評価方式は数多く提案されており，それぞれが目的に応じた利点を持つ．本コンテストでは 3.3, 3.4 節で定義した指標を採用し，公平に評価する．だが，これによって匿名加工処理の有益性・安全性のすべてを網羅できるわけではない．本コンテストの趣旨は多くの匿名加工手法と再識別手法を掛け合わせ，それぞれ手法の利点/問題点を検討することにある．そのため，評価方式は絶対的な指標の大小ではなく，参加者同士の相対評価である順位 $Rank$ を主に用いる．

総合指標の問題点は準備段階でも多く判明しており，3.5 節山岡匿名加工や，データ量を極端に少なくする不正行為など，参加者が指標の問題点をついた対応を行わないよう，ルール上の禁止事項を設けたうえで予備戦と本戦を分けることで対応する．

匿名加工コンテストでは，最も有用性が高く，最も安全な匿名加工者を勝者とする．匿名加工データ $\boldsymbol{Y}$ の総合スコア $\text{Score}(\boldsymbol{Y})$ は有用性についての評価値

$$\frac{1}{6}\sum_{i=1}^{6} Rank(U_i)$$

と安全性についての評価値

$$\frac{Rank(S_1) + Rank(S_2)}{4} + \frac{1}{2} Rank(\max_j \text{re-id}^{E_j})$$

の和で定める．ここで，**表 5** に示す有用性指標 $U_1, \ldots, U_6$，および安全性指標 $S_1, S_2, E_1 \ldots, E_4$ を用いる．

匿名加工データの安全性は，いかなる攻撃に対しても最低限保証される安全性と解釈し，複数の再識別アルゴリズムの中の最も高い再識別率で評価する．これに対して，再

表 6 上位 10 匿名加工データ
Table 6 Top 10 Anonymized data.

team ID	U_1	U_2	U_3	U_4	U_5	U_6	S_1	S_2	E_1	E_2	E_3	E_4	Max E_i
02	0.00	0.00	0.00	0.09	0.01	0.00	1.00	13.71	0.00	0.00	0.00	0.00	0.00
02	0.00	0.00	0.00	0.09	0.01	0.00	1.00	13.68	0.00	0.00	0.00	0.00	0.00
01	0.00	0.00	0.00	0.07	0.02	0.00	1.00	36.07	0.00	0.02	0.00	0.00	0.02
02	0.00	4321.75	1.54	0.03	0.01	0.00	3.00	36.07	0.00	0.02	0.01	0.00	0.02
10	0.00	0.00	0.00	0.00	0.03	0.00	1.00	36.07	0.00	0.08	0.08	0.01	0.08
15	0.00	31400.95	0.99	0.00	0.02	0.00	3.00	4.86	0.19	0.24	0.25	0.05	0.25
07	0.00	46944.41	2.16	0.00	0.02	0.00	5.00	89.60	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.03	0.00	1.00	36.07	0.00	0.07	0.07	0.01	0.07
10	0.00	0.00	0.00	0.00	0.03	0.00	1.00	36.07	0.00	0.07	0.07	0.01	0.07
15	0.00	31572.91	1.01	0.00	0.02	0.00	3.00	4.91	0.20	0.24	0.25	0.05	0.25

表 5 有用性指標と安全性指標
Table 5 Utility and security measures.

No.	指標	概要	節
U_1	meanMAE	SA 平均絶対誤差	3.3.1
U_2	crossMean	クロス集計値の平均絶対誤差	3.3.2
U_3	crossCnt	クロス集計数の平均絶対誤差	3.3.2
U_4	corMAE	SA の相関係数の平均絶対誤差	3.3.3
U_5	IL	匿名加工データの各値の平均絶対誤差	3.6
U_6	nrow	匿名加工データのレコード数	3.3.4
S_1	k-anony	k-匿名性指標の最小値	3.4.1
S_2	k-anonyMean	k-匿名性指標の平均値	3.4.1
E_1	IdRand	QI からランダムな再識別率	2.6
E_2	IdSA	QI から SA15 列による再識別率	2.6
E_3	Sort	SA の総和でソートによる再識別率	2.6
E_4	SA21	SA21 列について再識別率	2.6

識別アルゴリズムは，再識別に成功したレコード数の相和を採用する．再識別攻撃者の優劣は，データの種類にかかわらず，最も多くのユーザを識別した量で評価されるべきと考える．

これらの指標を用いて予備戦を行い，参加者の評価分布や傾向に応じて指標を再検討し，本戦では，対象データの分布変更，指標の重み付け，評価対象列の変更などを実施して公平性を高める．

4. コンテスト結果

4.1 匿名加工データ

表 6 と**表 7** に本コンテストの上位 10 位の匿名加工データとオリジナルデータの，安全性，有用性評価値を示す．チームは 3 個まで加工データを提出することが許されており，チーム 02 が上位の中に 3 つ入っている．

4.2 有用性の分布

図 4 に，有用性指標 U_1 の分布図を示す．ここで，X 軸はそれぞれの指標の値でソートされたランクであり，Y 軸は平均 0，分散 1 に正規化している．

匿名加工データごとに U_1，U_3，U_5 の指標を示した**図 5** によると，データについて有用性の値が異なっていることが分かる．すべての有用性を高めることは困難であるので，それぞれの戦略で優先する指標を選択していることが窺える．

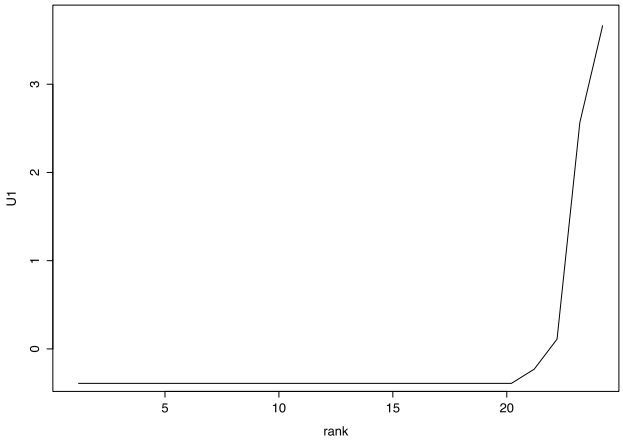

図 4 有用性 U_1 の分布
Fig. 4 Distribution of utility U_1.

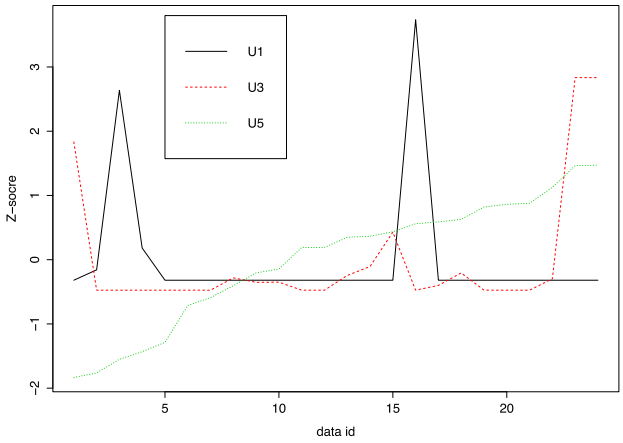

図 5 有用性 U_1，U_3，U_5 の関係
Fig. 5 Relationship among utilities U_1, U_3, U_5.

図 6 に有用性の代表としての U_5（X 軸）と安全性を示す再識別率の最大値（Y 軸）との散布図を示す．図の下に位置するデータ程，再識別率が低く，すなわち，安全性が高

表 7 オリジナル個人データ $\boldsymbol{X}$ の性質
Table 7 Measures of original personal data $\boldsymbol{X}$.

team ID	U_1	U_2	U_3	U_4	U_5	U_6	S_1	S_2	E_1	E_2	E_3	E_4	Max E_i
$\boldsymbol{X}$	0.00	0.00	0.00	0.09	0.01	0.00	1.00	1.87934145	0.65079	1.0	1.0	1.0	1.0

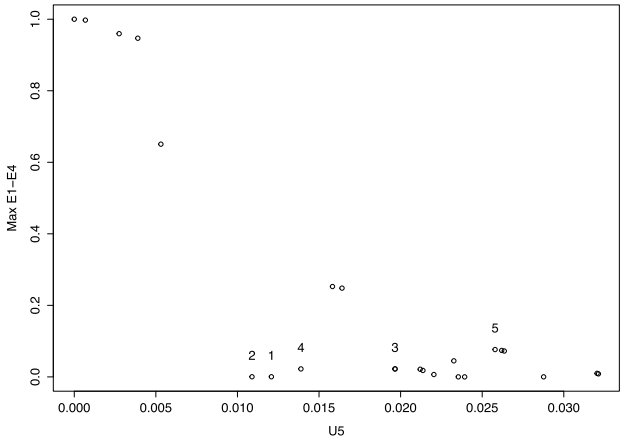

図 6 有用性と安全性の関係
Fig. 6 Tradeoff between utility and security.

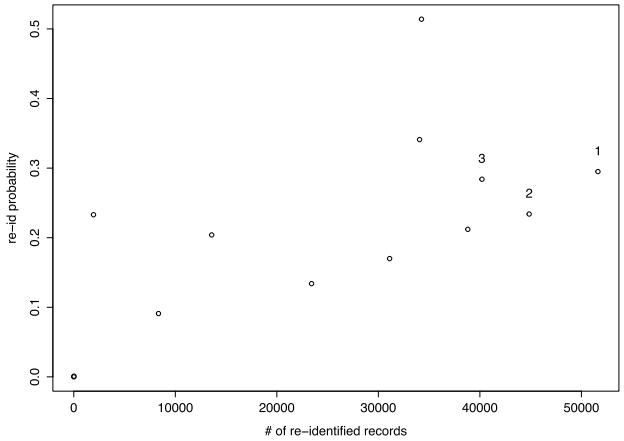

図 7 再識別総レコード数と再識別率
Fig. 7 Total number of re-identified records and re-id.

表 8 再識別率ランキング
Table 8 Re-identification ranking.

rank	# of identified records	trials	re-id	
1	51628	21	174993	29.5
2	44852	23	191659	23.4
3	40204	17	141661	28.4
4	38811	22	183326	21.2
5	34247	8	66664	51.4
6	34059	12	99996	34.1
7	31110	22	183326	17.0
8	23420	21	174993	13.4
9	13584	8	66664	20.4
10	8344	11	91663	9.1
11	1943	1	8333	23.3
12	18	2	16666	0.1
13	5	5	41665	0.0

い．右へ行くほど誤差が大きく，有用性が低い．したがって，最も左下に位置する匿名加工データが安全性と有用性の両方について良いデータである．

有用性を高めるためには，加工の幅を小さくする必要があり（左に向かう），それゆえ再識別も容易になる（上に向かう）．この散布図は，有用性と安全性の間に緩やかなトレードオフがあることを示している．ただし，ここで数字 1 から 5 で示された上位 5 個の加工データについていえば，そのトレードオフの中でも例外的に左下に突き出て分布している．これらには再識別を防止する巧みな加工が施されているためである．

4.3 再識別率の評価

表 8 に，再識別のチームごとの順位を示す．表では，順位，再識別に成功したレコード数，再識別を試みた加工データ数，再識別を試みた総レコード数，そして，再識別率を与えている．13 チームの平均再識別率は 20.9%である．

図 7 に再識別レコード数に対する再識別率の散布図を示す．数字は上位 3 位までのデータである．再識別の評価は，自チームを除く再識別合計のレコード数であり，この図の X 軸でのみ順位が決まる．たとえば，1 位のチームの再識別率は 29.4%であり，これは，5 位のチームの再識別率 51.4%より低い．識別率が低くても，より多くのレコードを識別したことが示されている．

5. 結論

個人情報データベースの匿名加工に関する社会的な整備状況に付いて解説し，その匿名加工と再識別の技術を競うコンテストを報告した．本コンテストで提案した技術の基本定義と有用性と安全性の評価指標は，考慮した不正行為と対象としたリスクを明らかにし，再識別の際に考慮しなくてはならないポイントを提供している．実装した評価プラットフォームは，参加者の評価を支援し，公平な安全性と有用性の定量的な評価が可能であることを立証した．

本コンテストが明らかにしたことは次のとおりである．

- どんなに巧妙に加工を加えても，技術を駆使すればデータの一部は再識別されてしまうことが分った．チーム平均識別率は 20.9%であり，完全な匿名加工は存在しないことを裏付けた．
- データの有用性にはユースケースに依存して多くの観点があり，単一の評価基準で計ることは難しい．コンテストで用いたように，$U_1, \ldots, U_6$ などの複数の指標を用いて総合的に評価することが必要である．

- 安全性と有用性の間にはトレードオフがある．コンテストに提出された匿名加工データは，有用性が高く安全性が低いグループと逆に安全性が高く有用性が低いグループ，そして，（上位 4 位のように）その両方とも高いグループの 3 つに分かれた．
- 上位にランクされた匿名加工は，山岡匿名加工（レコード順序の置換）を基本に複数の加工方法を組み合せたものが多かった．

コンテストで再識別に用いられたアルゴリズムは，チーム内の戦略や発見的なノウハウなどを含むため，コンテスト後も原則非公開である．また，本コンテストでは難易度が高いために履歴のような時系列データの加工は見送っていたが，今後はそれらに対してもコンテストの対象を広げる予定である．ここで得られた知見を，解析して方式を整理し，暗号技術のようにいずれは誰もが安心して用いることができる技術につなげて行きたい．

謝辞 本稿の執筆，ならびに，本コンテストの実施には，「擬似ミクロデータ（平成 16 年全国消費実態調査）」（独立行政法人統計センター）を利用した．また，本コンテストは，PWS 実行委員 CUP ワーキンググループの NTT ドコモ先進技術研究所山口高康氏，NTT セキュアプラットフォーム研究所濱田浩気氏，富士通研究所山岡裕司氏，ニフティ小栗秀暢氏，筑波大学佐久間淳氏の多大な協力の下で企画した．メンバーの熱心な討論と情熱に感謝いたします．

参考文献

[1] 菊池浩明，山口高康，濱田浩気，山岡裕司，小栗秀暢，佐久間淳：匿名加工・再識別コンテスト Ice & Fire の設計，コンピュータセキュリティシンポジウム（CSS 2015），プライバシーワークショップ，2B2-1, pp.1–8 (2015).
[2] Information Commissioner's Office (ICO): Anonymisation: Managing data protection risk code of practice (2012).
[3] El Emam, K. and Arbuckle, L.: Anonymizing Health Data Case Studies and Methods to Get You Started, O'Reilly (2013)（木村による和訳あり）.
[4] PWS CUP 2015 ウェブサイト，入手先 ⟨http://www.iwsec.org/pws/2015/pwscup.html⟩.
[5] Domingo-Ferrer, J. and Torra, V.: A quantitative comparison of disclosure control methods for microdata, Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies, pp.111–133 (2001).
[6] Domingo-Ferrer, J., Ricci, S. and Soria-Comas, J.: Disclosure Risk Assessment via Record Linkage by a Maximum-Knowledge Attacker, *2015 13th Annual Conference on Privacy, Security and Trust* (*PST*), IEEE (2015).
[7] Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Metayer, D.L., Tirtea, R. and Schinffer, S.: Privacy and Data Protection by Design – from policy to engineering, *ENISA* (2014).
[8] 秋山裕美，山口幸三，伊藤伸介，星野なおみ，後藤武彦：教育用擬似ミクロデータの開発とその利用—平成 16 年全国消費実態調査を例として，統計センター製表技術参考資料，16, pp.1–43 (2012).
[9] 大豆生田崇志：生易しくない匿名加工情報，日経コンピュータ，No.2015-7-23, pp.28–31 (2015).
[10] 佐藤慶浩：個人情報保護に関する国際動向：3) 国際規格動向，情報法制度研究会第 4 回シンポジウム，2016 入手先 ⟨http://www.dekyo.or.jp/kenkyukai/data/4th/20160612_Doc4-3.pdf⟩，（参照 2016-06）.

付　　録

A.1 簡易データの特性

表 A·1 に簡易データの特性を示す．属性 1 から 13 までは質的属性（i は整数，f は因子）で QI として扱われる．残りの属性 14 から 25 は実数を取る量的属性であり，SA とみなす．質的属性については，値の数とメジアンを，量的属性については平均値を示す．

3-匿名性を満足するために，世帯に関する属性 4，5，6，および，世帯主に関する属性 8，9，10，11，12，13 の優先順位で $k < 3$ となるレコードの値を一般化し，「不詳（V または VV）」で置き換えられている．

表 A·1 簡易データの特性
Table A·1 Attribute list of the NSTAC sudo microdata.

	属性	変数名	種類	値数	平均/メジアン	例
1	世帯区分	SetaiKubun	i	1	1	1（勤労）
2	人員	SetaiJinin	i	1	4	4
3	有業人員	ShuugyouJinin	i	1	1.504	1
4	住居構造	Kouzou	f	5	1	“1”（木造), “2”, “3”, “V”
5	建て方	Tatekata	f	7	1	“1”（一戸建), “2”, “3”, “V”
6	所有	Shoyuu	f	8	1	“1”（持家), “2”, “3”, “V”
7	性別	S1_Sex	i	1	1	1（男）
8	年齢	S1_Age	f	11	5	“1”（1-18 歳), “2”, “3”, “VV”
9	就業	S1_Shuugyou	f	3	1	“1”（就業), “2”, “V”
10	企業区分	S1_KigyouKubun	f	3	1	“1”（民営), “3”, “V”
11	企業規模	S1_KigyouKibo	f	7	3	“ ”（未回答), “1”, “2”, “3”, “V”
12	産業符号	S1_Sangyou	f	15	VV	“5”（建設業), “6”, “7”, “VV”
13	職業	S1_Shokugyou	f	6	1	“1”（常設作業者), “2”, “3”, “VV”
14	集計用乗数	Weight	n	–	15.741	13.2
15	消費支出	Youto037	n	–	324,525	155006
16	食糧	Youto038	n	–	74,639	25227
17	住居	Youto079	n	–	14,686	2000, 0
18	光熱	Youto084	n	–	19,733	18333
19	家具	Youto089	n	–	8,665	3834
20	衣類	Youto099	n	–	14,166	4784
21	医療	Youto117	n	–	11,104	5953
22	交通通信	Youto122	n	–	47,977	18767
23	教育	Youto129	n	–	33,928	29090
24	娯楽	Youto133	n	–	32,401	5514
25	その他	Youto142	n	–	67,227	20455

菊池 浩明 （正会員）

1988 年明治大学工学部電子通信工学科卒業．1990 年同大学大学院博士前期課程修了．1994 年同大学院博士（工学）．1990 年（株）富士通研究所入社．1994 年東海大学工学部電気工学科助手．1995 年同専任講師．1999 年同助教授，2006 年同情報理工学部情報メディア学科教授．1997 年カーネギーメロン大学計算機科学学部客員研究員．2013 年明治大学総合数理学部先端メディアサイエンス学科教授．2016 年同大学院先端数理科学研究科長．WIDE プロジェクト暗号メールシステム FJPEM の開発，認証実用化実験協議会（ICAT），IPA 独創情報技術育成事業等に従事．暗号プロトコル，ネットワークセキュリティ，ファジィ論理，プライバシー保護データマイニング等に興味を持つ．1990 年日本ファジィ学会奨励賞，1993 年情報処理学会奨励賞，1996 年 SCIS 論文賞，2010 年情報処理学会 JIP Outstanding Paper Award．2013 年 IEEE AINA Best Paper Award．2014 年情報セキュリティ文化賞．電子情報通信学会，日本知能情報ファジィ学会，IEEE，ACM 各会員．本会フェロー．