情報処理学会論文誌　数理モデル化と応用　Vol.11　No.3　100–105 (Dec. 2018)

データ間の距離に基づく情報損失指標

秋山 寛子[1,a)]　和田 昌昭[2]

受付日 2017年5月25日，再受付日 2017年7月14日 / 2018年1月27日 / 2018年4月30日，
採録日 2018年9月28日

概要：本論文では，データ間距離を用いてデータセット全体の持つ情報の量として情報容量を定義する．また，ミクロアグリゲーションにより失われる情報容量の割合を ILD と表し，これを新たな情報損失指標として提案する．ILD は数値データと非数値データの両方を含むデータセットに適用可能である．

キーワード：情報損失，ミクロアグリゲーション，匿名化

Information Loss Index Based on Distance between Data

Hiroko Akiyama[1,a)]　Masaaki Wada[2]

Received: May 25, 2017, Revised: July 14, 2017/January 27, 2018/April 30, 2018,
Accepted: September 28, 2018

Abstract: In this paper, we define information amount of a dataset based on distance between data. The ratio of the information amount lost by microaggregation is denoted ILD, and we propose ILD as a new index for measuring information loss. ILD can be applied to datasets consisting of both numeric and non-numeric attributes.

Keywords: information loss, microaggregation, anonymization

1. はじめに

データセットをグループに分割し，グループごとにデータを代表値に置き換える操作はミクロアグリゲーションとよばれ，匿名化の一手法として研究されている [1], [2]．あるデータセットにミクロアグリゲーションを実行すると，データセットの持つ情報の一部が失われる．このとき，失われる情報とはそもそも何なのか，一体どの程度の情報が失われたと考えればよいのか，というようなことが問題になる．

本論文では，データ間の距離が任意に与えられたとき，それに基づいてデータセット全体の持つ情報の量を定義し，それを情報容量とよぶ．また，ミクロアグリゲーションにより失われる情報容量の割合として情報損失指標 ILD (Information Loss based on Distance) を定義する．

ILD の特徴は，データに対する適切な距離さえ与えれば情報損失が自動的に得られるという汎用性である．カテゴリデータのような非数値データや，数値と非数値の属性が混在するデータセットにも柔軟に対応できることが特徴である．

数値データセットに対するミクロアグリゲーションにおいては，分割されたグループごとの代表値としてグループの平均値を用いる場合がよく研究されており，平均との差の 2 乗和に基づいた情報損失指標 ILSSDM (Information Loss based on Sum of Square Difference from the Mean) が用いられている [3], [4], [5], [6], [7], [8], [9]．データ間の距離としてユークリッド距離を用いた場合，本論文の ILD はこの ILSSDM と同じ値になり，その意味で ILD は ILSSDM の拡張となっている．

また，情報エントロピーを用いた情報損失も広く研究されている [10], [11], [12], [13]．情報エントロピーに関する情報損失は個別のデータがどの程度識別できなくなるかを測っているのに対し，ILD は 2 つのデータがどの程度区別

1 長野工業高等専門学校
National Institute of Technology, Nagano College, Nagano 381–8550, Japan
2 大阪大学大学院情報科学研究科
Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan
a) h_akiyama@nagano-nct.ac.jp

できなくなるかを測っている．k-匿名化のような応用分野では，2 つのデータどうしの区別可能性を評価したいため，ILD は有用な情報損失指標である．

2. 情報損失指標 ILD の定義

2.1 情報容量の定義

データセット

$$A = (x_1, \ldots, x_N), \quad x_i \in X \ (i = 1, \ldots, N)$$

があり，X の 2 つのデータ x，y 間の距離 $d(x, y)$ が与えられているとする．このとき，データセット A の情報容量 $I(A)$ を，

$$I(A) = \sum_{i=1}^{N} \sum_{j=1}^{N} d(x_i, x_j)^2$$

で定義する．

距離 $d(x, y)$ については，

(i) $d(x, y) \geq 0$ かつ $d(x, y) = 0 \Leftrightarrow x = y$

(ii) $d(x, y) = d(y, x)$

を仮定するが，三角不等式は成り立たなくても以下の議論に影響はない．具体的には次のようなものを考えている．

- ユークリッド距離
  $X = \mathbb{R}^n$，$x, y \in \mathbb{R}^n$ に対し，
  $$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$
- 離散距離
  X は任意の集合，$x, y \in X$ に対し，
  $$d(x, y) = \begin{cases} 1 & (x \neq y) \\ 0 & (x = y) \end{cases}$$
- 直積距離
  d_1，d_2 をそれぞれ X_1，X_2 上の距離とするとき直積集合 $X = X_1 \times X_2$ 上の距離が
  $$d((x_1, x_2), (y_1, y_2)) = \sqrt{w_1 d_1(x_1, y_1)^2 + w_2 d_2(x_2, y_2)^2}$$
  で定義される．重み w_1，w_2 は任意の正の実数である．

$I(A)$ は，データどうしの関連性をデータ間の距離によって与えたうえでデータセット A に含まれる情報の量を測るものである．情報エントロピーと区別するために，情報量とはよばずに情報容量とよんでいる．

2.2 ILD の定義

データセット

$$A = (x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, \ldots, x_{m1}, \ldots, x_{mn_m})$$

がグループ $A_1, \ldots, A_m$ に分割されているとする．ただし，

$$A_k = (x_{k1}, \ldots, x_{kn_k}) \ (k = 1, \ldots, m)$$

である．いま，グループ A_k に属するデータをすべて A_k の代表値 $\hat{x}_k$ に置き換えると，データセットは，

$$\widehat{A} = (\hat{x}_1, \ldots, \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_2, \ldots, \hat{x}_m, \ldots, \hat{x}_m)$$

となる．このとき，一般に $I(A) > I(\widehat{A})$ となっている．

上記の置き換えに関する情報損失を，

$$ILD = \frac{I(A) - I(\widehat{A})}{I(A)}$$

と定義する．

以下に，ILD の具体例を示す．

例 2.1（数値データセット）　データセット $A = (1, 2, 3, 4)$ が $A_1 = (1, 2)$ と $A_2 = (3, 4)$ に分割されているとする．グループ A_1，A_2 に属するデータをそれぞれのグループの平均値に置き換えると，データセットは $\widehat{A} = (1.5, 1.5, 3.5, 3.5)$ となる．データ間の距離をユークリッド距離で与えるとすると，A と $\widehat{A}$ の情報容量はそれぞれ，$I(A) = 40$，$I(\widehat{A}) = 32$ となる．したがって，平均値に置き換えることによる情報損失は，$ILD = \frac{40-32}{40} = 0.2$ である．

例 2.2（数値と記号の組）　数値と記号の組からなるデータセット $A = ((1, a), (2, a), (3, b), (4, c))$ が $A_1 = ((1, a), (2, a))$ と $A_2 = ((3, b), (4, c))$ に分割されているとする．グループ A_1，A_2 に属するデータをそれぞれ $(1.5, a)$ と $(3.5, b)$ に置き換えると，データセットは $\widehat{A} = ((1.5, a), (1.5, a), (3.5, b), (3.5, b))$ となる．数値の距離をユークリッド距離，記号の距離を $d(a, b) = 1$，$d(b, c) = 1$，$d(a, c) = 3$ とし，直積距離を用いてデータ間の距離を計る．数値データと記号データを対等に評価するために重みをそれぞれ数値データのみの情報容量，記号のみの情報容量の逆数として $w_1 = 1/40$，$w_2 = 1/42$ とすると，$I(A) = 2$，$I(\widehat{A}) = 104/105$ となる．したがって，この場合の情報損失は $ILD = \frac{2-104/105}{2} = \frac{53}{105}$ である．

3. 情報損失指標の比較

本章では，数値データセットに対して定義された情報損失指標や，情報エントロピーを用いた情報損失，非数値データセットに対して定義された情報損失指標と ILD の比較について述べる．

3.1 ILSSDM との比較

数値データセットに対して，分割されたグループごとの代表値をグループの平均値とし，グループ内のデータを平均値に置き換えることを平均化とよぶ．本節では，数値データセットに対して平均化を行う場合の ILD が，ILSSDM と同じになることを示す．また，ユークリッド空間に埋め込み不可能な距離を持つデータセットについての考察を述べる．

まず，ILSSDM の定義を示す．N 個のデータからなる数値データセット $X \in \mathbb{R}^n$ がグループ $X_1, \ldots, X_m$ に分割され，X_i が n_i 個のデータ x_{ij}（$0 \leq j \leq n_i$）を含んでいるものとする．X_i の平均を $\overline{x}_i$ とし，平均からの距離の 2 乗和（Sum of Squared Errors）を

$$SSE = \sum_{i=1}^{m}\sum_{j=1}^{n_i} |x_{ij} - \overline{x_i}|^2$$

で表す．一方，全データの平均を $\overline{x}$ とし，

$$SSA = \sum_{i=1}^{m} n_i |\overline{x_i} - \overline{x}|^2$$

とおけば，全データの平均からの距離の 2 乗和は

$$SST = \sum_{i=1}^{m}\sum_{j=1}^{n_i} |x_{ij} - \overline{x}|^2 = SSE + SSA$$

と表すことができる．このとき，ILSSDM は次のように定義される．

$$ILSSDM = \frac{SST - SSA}{SST} = \frac{SSE}{SST}$$

これは，X_i の分散 $var(X_i)$ と X の分散 $var(X)$ を用いれば，

$$ILSSDM = \frac{\sum_{i=1}^{m} n_i\, var(X_i)}{N\, var(X)} \tag{1}$$

と表すことができる．

次に，同様の分割における情報容量を示す．同様の分割において，全データの情報容量は，

$$\begin{aligned} I(X) & \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{x \in X_i}\sum_{y \in X_j} |x - y|^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{x \in X_i}\sum_{y \in X_j} |(x - \overline{x_i}) + (\overline{x_i} - \overline{x_j}) + (\overline{x_j} - y)|^2 \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{x \in X_i}\sum_{y \in X_j} (|x - \overline{x_i}|^2 + |\overline{x_i} - \overline{x_j}|^2 + |\overline{x_j} - y|^2) \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j (var(x_i) + |\overline{x_i} - \overline{x_j}|^2 + var(x_j)) \end{aligned} \tag{2}$$

となる．また，X の平均 $\overline{x}$ を用いた計算によって，

$$\begin{aligned} I(X) &= \sum_{x \in X}\sum_{y \in X} |x - y|^2 \\ &= \sum_{x \in X}\sum_{y \in X} |(x - \overline{x}) + (\overline{x} - y)|^2 \\ &= \sum_{x \in X}\sum_{y \in X} (|x - \overline{x}|^2 + |\overline{x} - y|^2) \\ &= 2N^2\, var(X) \end{aligned} \tag{3}$$

と表すこともできる．

最後に，平均化に関する ILD と ILSSDM を比較する．X の分割 $X_1, \ldots, X_m$ において，各 X_i に含まれるデータをグループの平均値 $\overline{x}_i$ に置き換えたデータセットを X' とすると，

$$I(X') = \sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j |\overline{x}_i - \overline{x}_j|^2 \tag{4}$$

であるから，(2)，(4) より，

$$\begin{aligned} I(X) - I(X') &= \sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j (var(X_i) + var(X_j)) \\ &= 2\sum_{i=1}^{m}\sum_{j=1}^{m} n_i n_j\, var(X_i) \\ &= 2N \sum_{i=1}^{m} n_i\, var(X_i) \end{aligned} \tag{5}$$

したがって，(3)，(5) より，

$$\begin{aligned} ILD &= \frac{I(X) - I(X')}{I(X)} \\ &= \frac{\sum_{i=1}^{m} n_i\, var(X_i)}{N\, var(X)} \end{aligned} \tag{6}$$

となり，(1)，(6) より，ILD は ILSSDM に一致することが分かる．したがって，ILD は ILSSDM の拡張となっている．

ILSSDM と比較した場合の ILD の特徴は，ユークリッド空間でなくても計算が可能である点である．数値属性と非数値属性を含むデータセットについて，データ間距離として直積距離を設定した場合，非数値データセットをデータ間距離を変えずにユークリッド空間に埋め込めるとき，ILSSDM と ILD は同じ値になる．しかし，2.2 節の例 2.2 のようにユークリッド空間に埋め込めない距離を持つデータセットも存在し，その場合は ILSSDM によって情報損失を評価することができない．

3.2　情報エントロピーとの比較

本節では，情報エントロピーに関する情報損失と ILD を比較する．

データ間距離として離散距離を設定した場合，情報容量は値が異なる 2 つのデータの組合せを数えあげているものであり，この場合の ILD は 2 つのデータが区別できなくなる割合を表しており，匿名化の効果を表す 1 つの指標として用いることができる．

例として，$N = nk$ 個の異なるデータがあり，それらを k 個ずつ n 個のグループに分割して，グループ内のデータをグループの代表値に置き換える匿名化を考える．グループの数 n を固定して k が十分大きい場合を考えると，最初はどの 2 つも区別可能であったデータが，匿名化後は同じ

グループに属する場合に区別不可能になるのであるから，その割合は約 $1/n$ である．離散距離に対する ILD は，この割合を測っているものである．実際，匿名化前の情報容量は，

$$I(X) = N(N-1)$$

で，分割後の情報容量は

$$I(\widehat{X}) = k^2 \cdot n(n-1) = N(N-k)$$

であるから，

$$ILD = \frac{k-1}{N-1} = \frac{k-1}{nk-1}$$

となり，k が大きいとき ILD $\sim 1/n$ である．

一方，カテゴリデータなどの非数値データについては，情報エントロピーに関する情報損失もよく研究されている．上の例において，データの生起確率はすべて $1/N$ であるものとして考えると，匿名化前の情報エントロピーは，

$$H(X) = \log N = \log n + \log k$$

で，匿名化後の情報エントロピーは

$$H(\widehat{X}) = \log n$$

であるから，情報エントロピーに関する情報損失は

$$\frac{H(X) - H(\widehat{X})}{H(X)} = \frac{\log k}{\log N} = \frac{\log k}{\log n + \log k}$$

となる．グループの数 n を固定して考えると，情報エントロピーに関する情報損失は，k が大きくなるに従いゆっくり 1 に近づく．これは，データセットに含まれる個々のデータの識別情報が徐々に失われていくことを表している．

このように，離散距離に関する ILD は 2 つのデータがどれぐらい区別できなくなるかを測るのに対し，情報エントロピーに関する情報損失は個別のデータがどれぐらい識別できなくなるかを測っており，情報損失指標として性質が大きく異なっている．匿名化のような応用分野では，個々のデータが識別，特定できなくなる度合いは重要な評価項目である．k-匿名化は，データセットに含まれるどのデータも k 人と区別がつかないようにデータセットを加工する手法であり，その評価においては，識別可能性に加えて区別可能性も考慮する必要がある．したがって，ILD は k-匿名化に対して有効な情報損失指標である．

3.3　非数値データの情報損失指標

非数値データの匿名化に対して定義された情報損失指標について述べる．

Classification Metric（CM）[14] は，カテゴリデータをグループに分割する一般化という手法とデータの一部を削除する手法を用いた匿名化に適用可能である．データセットの各データには，クラスというラベルが付与されるとする．グループに分割された状態において，各グループに含まれるデータのうち，多数を占めるクラスをマジョリティ，マジョリティでないクラスをマイノリティとする．CM は，マイノリティのデータと削除されたデータを数えあげデータ総数で割ったものである．複数の属性を持つデータセットへも適用可能であるが，カテゴリデータの組に対してラベルが付与されるため，連続値を持つ数値属性とカテゴリ属性が両方含まれるデータセットには適用できない．

Discernibility Metric（DM）[15] は，一般化と削除を用いた匿名化の結果について，データの識別可能性に基づき情報損失を表す指標であり，グループに含まれるデータ数を用いて計算される．たとえば，異なる 5 つの値のデータからなるデータセットを，データ数が 2 つと 3 つのグループに分割する場合，どのデータを同じグループにしても DM の値は同じになる．同じグループに含まれるデータどうしの類似度は考慮されていないため，元のデータと比較した情報損失を表していない．

4.　数値と非数値の属性を含むデータセットへの ILD の適用

本章では，数値属性と非数値属性が両方含まれるデータセットを匿名化し，その結果について ILD を計算し，ILD の適用について考察する．

4.1　ILD を適用する実験データセット

ILD を適用するデータセットとして，アメリカの国勢調査の結果を抽出した UCI adult dataset [16] を使用する．このデータセットのデータ総数は 32,561 である．また，含まれる属性数は 14 で，そのうち数値属性（整数値）が 5 個，非数属性（カテゴリデータ）が 9 個である．本実験では，数値データの属性である capital gain と，カテゴリデータの属性である marital status を取り出し，2 属性のデータ組のデータセットに対して ILD を計算する．capital gain は 0 から 99,999 の整数値をとり，marital status は 7 つのカテゴリデータのいずれかの値をとる．

4.2　データセットのミクロアグリゲーションの方法

データセットに対して異なる分割を行い，3 種類の分割されたデータセットを生成する．

1 つ目の分割は，数値データに関してより近い値どうしでグループをつくる分割である．データセットを数値データについて昇順になるようにソートし，小さい値から順に k 個ずつのグループをつくり，残りのデータが $2k-1$ 個以下になったらそれらを 1 つのグループとする．このようにして分割されたデータセットを D_A とする．

2 つ目の分割は，カテゴリデータに関してより近い値ど

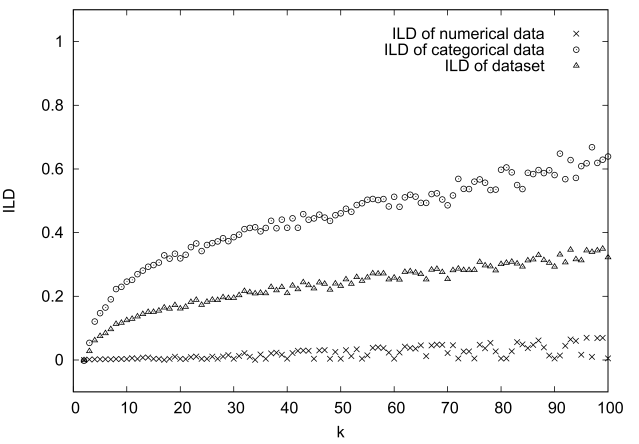

図 1　D_A の ILD
Fig. 1　ILD for D_A.

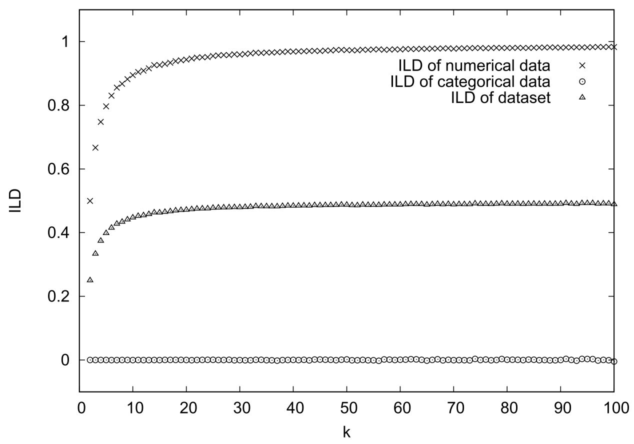

図 2　D_B の ILD
Fig. 2　ILD for D_B

うしでグループをつくる分割である．カテゴリデータの文字列について昇順になるようにソートし，同様に k 個ずつのグループをつくる．このようにして分割されたデータセットを D_B とする．

3 つ目の分割は，どちらの属性についてもより近い値どうしでグループをつくる分割である．まず，カテゴリデータの文字列について昇順になるようにソートし，次に同じカテゴリデータを持つデータを数値について昇順になるようにソートする．そして同様に k 個ずつのグループをつくる．このようにして分割されたデータセットを D_C とする．

同じグループに含まれるデータを置き換える代表値は，数値データはグループの平均値，カテゴリデータはグループ内で現れる頻度が最も高いデータとする．

4.3　ILD の適用結果と考察

D_A，D_B，D_C について，数値属性のみの ILD，カテゴリ属性のみの ILD，データセット全体の ILD を計算し，その結果を**図 1**，**図 2**，**図 3** に示す．数値データのデータ間距離はユークリッド距離，カテゴリデータのデータ間距離は離散距離，データ組間の距離はユークリッド距離と離散距離の直積距離とした．また，直積距離の重みは，それぞれの属性の情報容量の逆数とした．

D_A は，数値データについて近い値どうしを同じグループにしたため数値属性の ILD は小さいが，カテゴリデータについては分割の際に値が考慮されていないため，同じグループに含まれるカテゴリデータの値はばらばらとなり，k が大きくなるにつれてカテゴリ属性の ILD は大きくなる．このため，データセット全体の ILD は，k が大きくなるにつれて大きくなる．D_B についても同様の結果となっている．両方の属性の値を考慮して分割した D_C は，いずれの属性の ILD も大きく増加しないため，データセット全体の ILD は D_A，D_B と比較して小さくなっている．

ILD の適用により，数値属性と非数値属性の両方を含むデータセットにも，情報損失を求めることが可能となった．

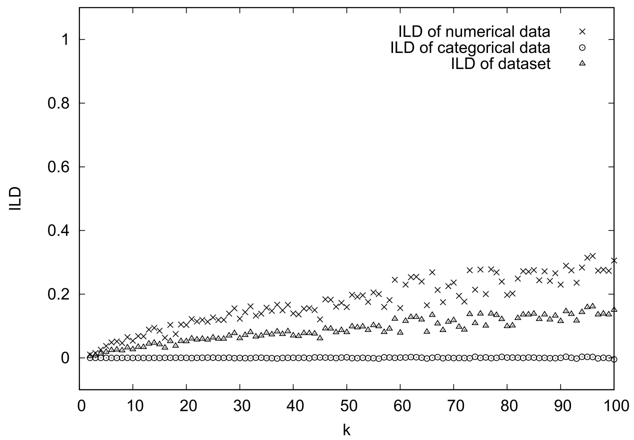

図 3　D_C の ILD
Fig. 3　ILD for D_C.

さらに，匿名化されたデータセットは，元のデータと比較してどの程度情報を損失したかを表すことができるようになった．これにより，より似た値どうしが同じグループに含まれているかというような，情報損失に関するミクロアグリゲーションの評価が可能となった．

5.　まとめ

本論文では，データセットにミクロアグリゲーションを行う際に生じる情報の減少を客観的に示す指標として，情報損失指標 ILD を提案した．ILD の定義のため，データセットの持つ情報の量を表す情報容量を定義した．これは，データ間の距離に基づき計算できるものである．情報容量を定義することにより，元のデータセットに対して，ミクロアグリゲーションにより変化したデータセットはどの程度情報が減少したかを表すことが可能となった．ILD は，数値属性と非数値属性の両方を含むデータセットへも適用でき，匿名化の評価に広く活用することができる．

謝辞　研究を進展させる有益な助言に対し本論文の査読者に感謝の意を表す．

参考文献

[1] Domingo-Ferrer, J. and Torra, V.: Ordinal, continuous and heterogeneous k-anonymity through microaggregation, *Data Mining and Knowledge Discovery*, Vol.11, No.2, pp.195–212 (2005).

[2] Domingo-Ferrer, J., Martínez-Ballesté, A., Mateo-Sanz, J.M. and Sebé, F.: Efficient multivariate data-oriented microaggregation, *The VLDB Journal—The International Journal on Very Large Data Bases*, Vol.15, No.4, pp.355–369 (2006).

[3] Edwards, A.W.F. and Cavalli-Sforza, L.L.: A method for cluster analysis, *Biometrics*, pp.362–375 (1965).

[4] Gordon, A.D. and Henderson, J.T.: An algorithm for Euclidean sum of squares classification, *Biometrics*, pp.355–362 (1977).

[5] Hansen, P., Jaumard, B. and Mladenovic, N.: Minimum sum of squares clustering in a low dimensional space, *Journal of Classification*, Vol.15, No.1, pp.37–55 (1998).

[6] MacQueen, J. et al.: Some methods for classification and analysis of multivariate observations, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, pp.281–297 (1967).

[7] Solanas, A., Martinez-Balleste, A. and Domingo-Ferrer, J.: V-mdav: A multivariate microaggregation with variable group size, *17th COMPSTAT Symposium of the IASC*, pp.917–925 (2006).

[8] Oganian, A. and Domingo-Ferrer, J.: On the complexity of optimal microaggregation for statistical disclosure control, *Statistical Journal of the United Nations Economic Commission for Europe*, Vol.18, No.4, pp.345–353 (2001).

[9] Domingo-Ferrer, J. and Mateo-Sanz, J.M.: Practical data-oriented microaggregation for statistical disclosure control, *IEEE Trans. Knowledge and Data Engineering*, Vol.14, No.1, pp.189–201 (2001).

[10] Domingo-Ferrer, J. and Rebollo-Monedero, D.: Measuring risk and utility of anonymized data using information theory, *Proc. 2009 EDBT/ICDT Workshops*, ACM (2009).

[11] De Waal, A.G. and Willenborg, L.C.R.J.: Information loss through global recoding and local suppression, *Netherlands Official Statistics*, Vol.14, pp.17–20 (1999).

[12] Kooiman, P., Willenborg, L. and Gouweleeuw, J.: PRAM: A Method for Disclosure Limitation of Microdata, Research Paper, No.9705, Statistics Netherlands, Voorburg (1998).

[13] Domingo-Ferrer, J. and Torra, V.: Disclosure control methods and information loss for microdata, Confidentiality, disclosure, and data access: Theory and practical applications for statistical agencies, pp.91–110 (2001).

[14] Iyengar, V.S.: Transforming data to satisfy privacy constraints, *Proc. 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM (2002).

[15] Bayardo, R.J. and Agrawal, R.: Data privacy through optimal k-anonymization, *Proc. 21st International Conference on Data Engineering, ICDE 2005*, IEEE (2005).

[16] Lichman, M.: UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences, available from ⟨http://archive.ics.uci.edu/ml⟩.

秋山 寛子　（正会員）

2015 年慶應義塾大学大学院メディアデザイン研究科単位取得退学．現在，長野工業高等専門学校電子情報工学科助教．電子情報通信学会会員．

和田 昌昭

1986 年コロンビア大学 Ph.D. 1987 年ペンシルバニア大学アシスタントプロフェッサー．1991 年ケース・ウェスタンリザーブ大学客員アシスタントプロフェッサー．1992 年から奈良女子大学講師，助教授，教授を経て，2009 年より大阪大学大学院情報科学研究科情報基礎数学専攻教授．バイオイメージング学会会員．