

k -匿名化アルゴリズムにおける情報損失の極小化

秋山 寛子^{1,a)} 和田 昌昭² 中山 雅哉³ 加藤 朗⁴ 砂原 秀樹⁴

受付日 2016年3月10日, 採録日 2016年9月6日

概要: プライバシ保護の観点から匿名化技術の重要度が高まっており, なかでも k -匿名化が活発に研究されている. k -匿名化のためのアルゴリズムとしては MDAV や VMDAV などが提案されているが, 数値データの匿名化において, それらは情報損失の観点から必ずしも最良のものとはなっていない. 本論文では, それらのアルゴリズムによって得られたデータの分割を, k -匿名性を保ったまま修正して情報損失を極小にする方法を提案する. またそれを実装して, いくつかのデータセットに適用し評価を行う. データ総数を k で割った余りが大きい場合の多くで, MDAV や VMDAV による情報損失を提案アルゴリズムにより改善可能である. また, 提案アルゴリズムは, データ総数によらず高速に実行可能である.

キーワード: 情報セキュリティ, プライバシ保護, 匿名化技術, k -匿名化, 情報損失

Minimizing Information Loss in Anonymization Algorithms

HIROKO AKIYAMA^{1,a)} MASAOKI WADA² MASAYA NAKAYAMA³ AKIRA KATO⁴ HIDEKI SUNAHARA⁴

Received: March 10, 2016, Accepted: September 6, 2016

Abstract: Anonymization is becoming more and more important for the purpose of privacy protection, and k -anonymization in particular is actively studied. Various k -anonymization algorithms such as MDAV and VMDAV are proposed, but they are not necessarily optimal for numerical data from the viewpoint of information loss. In this paper, we propose a method for minimizing information loss, while preserving k -anonymity, of the partition of dataset obtained by these algorithms. We implement the algorithm, apply it to several datasets and evaluate the results. In many cases where the remainder after dividing the number of data by k is large, the information loss of MDAV and VMDAV can be reduced by our algorithm. Also, our algorithm can be executed quickly regardless of the number of data.

Keywords: information security, privacy protection, anonymization technique, k -anonymity, information loss

1. はじめに

パーソナル情報の活用においてプライバシ保護の観点か

ら匿名化の重要性が増している. 匿名化指標の1つとして, k 人以上のうち誰のデータか区別がつかない状態を表す k -匿名性 (k -anonymity) がある [1]. データセットをいくつかのクラスタに分割し, クラスタに含まれるデータを平均値のような代表値に置き換えてしまうことによりクラスタ内のデータが互いに区別できないようにするのであるが, その際, どのクラスタも k 個以上のデータを含むならば分割は k -匿名性を持つといい, k -匿名性を持つようにデータセットを分割することを k -匿名化と呼ぶ.

匿名化によってデータが持つ情報の一部が失われることは避けられないが, データ利用者にとっては, なるべく値が近いデータどうしが同じクラスタにまとめられていて情報の損失が小さいことが望ましい. そのための k -匿名化ア

¹ 長野工業高等専門学校
National Institute of Technology, Nagano College, Tokuma,
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
³ 東京大学情報基盤センター
Information Technology Center, The University of Tokyo,
Bunkyo, Tokyo 113-8658, Japan
⁴ 慶應義塾大学大学院メディアデザイン研究科
Graduate School of Media Design, Keio University,
Yokohama, Kanagawa 223-8526, Japan
a) h.akiyama@nagano-nct.ac.jp

ルゴリズムとして MDAV [2] や VMDAV [3] が考案されている。MDAV は、アメリカの国勢調査や住宅調査などにも適用され実効性が確かめられている有名なアルゴリズムであり、VMDAV は、データの偏りを考慮して MDAV を改良したアルゴリズムである。

本論文では、属性数が 1 の数値データの匿名化を扱う。数値データをクラスタの平均値に置き換える匿名化によってどの程度情報が失われたかを表す指標として情報損失 (information loss) が研究されている [4], [5], [6], [7], [8]。MDAV や VMDAV は、この情報損失に関して必ずしも最良のものとはなっていない。本論文では、 k -匿名化によって得られた分割に対し、 k -匿名性を保ったまま分割を変更して情報損失を極小にする方法を提案し、それを具体的なデータに適用した結果を示す。MDAV や VMDAV による匿名化がどの程度改良可能かを調べ、また、提案アルゴリズムの評価も行う。

2. 情報損失

数値データセットがクラスタ D_1, \dots, D_g に分割され、 D_i が n_i ($\geq k$) 個のデータ x_{i1}, \dots, x_{in_i} を含んでいるものとする。 D_i の平均を \bar{x}_i とし、平均からの誤差 2 乗和 (sum of squared errors) を

$$SSE_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

とする。それらをすべてのクラスタについて加えた量を

$$SSE = \sum_{i=1}^g SSE_i$$

で表す。一方、全データの平均を \bar{x} とし、

$$SSA = \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})^2$$

とおけば、全データの \bar{x} からの誤差 2 乗和は

$$SST = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = SSE + SSA$$

と表すことができる。 SST が全データの持つ情報量、 SSA が匿名化されたデータの情報量と考えると、 SSE が匿名化によって失われた情報量を表している。このとき、情報損失は

$$\frac{SSE}{SST}$$

と定義される [3]。

3. 情報損失極小化アルゴリズム

本論文で提案するアルゴリズムは、 k -匿名性を持つ数値データセットの分割において、あるクラスタのデータ

を別のクラスタに移動することにより全体の情報損失を減少させようとするものである。数値データに MDAV や VMDAV を適用して得られる分割 D_1, \dots, D_g については、

$$D_i \text{ の最大データ} \leq D_{i+1} \text{ の最小データ} \quad (1)$$

という条件がすべての i で成り立つ。この条件を満たす分割が与えられているものとして話を進める。

3.1 データ移動の判定条件

あるクラスタのデータを x_1, \dots, x_n , 平均を \bar{x} とし、誤差 2 乗和を

$$SSE_x = \sum_{j=1}^n (x_j - \bar{x})^2$$

とする。このとき、任意の a に対し

$$\sum_{j=1}^n (x_j - a)^2 = SSE_x + n(\bar{x} - a)^2 \quad (2)$$

が成り立つことに注意する。このクラスタに新しいデータ x を追加したとき、 x_1, \dots, x_n, x の平均 \bar{x}' は

$$\bar{x}' = \frac{1}{n+1} (n\bar{x} + x) \quad (3)$$

となり、誤差 2 乗和 SSE'_x は式 (2) を用いて

$$\begin{aligned} SSE'_x &= \sum_{j=1}^n (x_j - \bar{x}')^2 + (x - \bar{x}')^2 \\ &= SSE_x + n(\bar{x} - \bar{x}')^2 + (x - \bar{x}')^2 \end{aligned} \quad (4)$$

であるから、式 (3), (4) より、

$$\begin{aligned} SSE'_x - SSE_x &= \frac{n+1}{n} (x - \bar{x}')^2 \\ &= \frac{n}{n+1} (x - \bar{x})^2 \end{aligned} \quad (5)$$

となる。

いま、クラスタ x_1, \dots, x_n, x と y_1, \dots, y_m があり、第 1 クラスタのデータ x を第 2 クラスタに移すことを考える。このとき、誤差 2 乗和の合計は、 $SSE'_x + SSE_y$ から、

$$\begin{aligned} SSE_x + SSE'_y &= \left(SSE'_x - \frac{n+1}{n} (x - \bar{x}')^2 \right) \\ &\quad + \left(SSE_y + \frac{m}{m+1} (x - \bar{y})^2 \right) \end{aligned}$$

に変化する。すなわち、

$$-\frac{n+1}{n} (x - \bar{x}')^2 + \frac{m}{m+1} (x - \bar{y})^2 < 0 \quad (6)$$

であれば、全体の SSE が減少する。したがって、式 (6) が成り立つ場合にデータの移動を行うという操作を繰り返すことにより情報損失が極小となる匿名化情報を生成することが可能となる。

3.2 アルゴリズムの概要

k -匿名性を満たす分割 D_1, \dots, D_g において条件 (1) が成り立っているものとする。このとき、 D_i, D_{i+1} の間でデータを移動することを考える。

(a) D_i, D_{i+1} のデータを

$$D_i = \{x_1, \dots, x_n, x\}$$

$$D_{i+1} = \{y_1, \dots, y_m\}$$

と表すこととし、 D_i の最大が x 、平均が \bar{x}' 、 D_{i+1} の平均が \bar{y} とする。このとき、 $n(D_i) = n + 1 > k$ かつ

$$-\frac{n+1}{n}(x - \bar{x}')^2 + \frac{m}{m+1}(x - \bar{y})^2 < 0 \quad (7)$$

ならば、データを移動したほうが情報損失が小さくなるため、 x を D_i から D_{i+1} に移動する。この操作を条件 (7) が成り立つ限り繰り返す。

(b) 条件 (7) が成り立たない場合、

$$D_i = \{x_1, \dots, x_n\}$$

$$D_{i+1} = \{x, y_1, \dots, y_m\}$$

とし、 D_i の平均が \bar{x} 、 D_{i+1} の最小が x 、平均が \bar{y}' とする。このとき、 $n(D_{i+1}) = m + 1 > k$ かつ

$$-\frac{n}{n+1}(x - \bar{x})^2 + \frac{m+1}{m}(x - \bar{y}')^2 > 0 \quad (8)$$

ならば x を D_{i+1} から D_i に移動する。この操作を条件 (8) が成り立つ限り繰り返す。

(a), (b) の処理をすべての隣り合うクラスタ D_i, D_{i+1} ($i = 1, \dots, g-1$) 間で行う。データの移動が1度でも発生した場合、再び先頭のクラスタから同様の処理を繰り返す。どの隣り合うクラスタについてもデータの移動がなかった場合、アルゴリズムを終了する。上記のアルゴリズムを MIL (minimizing information loss) と呼ぶ。アルゴリズムの流れを Algorithm 1 に示す。ここで、 $n(D_i)$ はクラスタ D_i のデータ数を表す。

データが昇順にソートされており、データとは別に各クラスタの要素数と平均を保持するものとすれば、条件 (7), (8) の判定と、データ移動があった場合の要素数や平均の更新はデータ数によらず一定時間内に効率良く行えることに注意する。

最後に、このアルゴリズムによって得られた分割は、1つのクラスタの要素を別のクラスタに移すことによってこれ以上情報損失を減少させることができないという意味で極小ではあるが、2個以上の要素の同時移動やクラスタ数の変更によりさらに情報損失が減少する可能性があるため、必ずしも情報損失最小とは限らないことを述べておく。

4. 実験

本章では、既存の k -匿名化アルゴリズムと提案アルゴリ

Algorithm 1 MIL (minimizing information loss)

```

flag ← 1
while flag = 1 do
    flag ← 0
    for i = 1 to g - 1 do
        while n(Di) > k do
            n = n(Di) - 1
            m = n(Di+1)
            X = - $\frac{n+1}{n}(x - \bar{x}')^2 + \frac{m}{m+1}(x - \bar{y})^2$ 
            if X < 0 then
                Move x from Di to Di+1
                flag ← 1
            else
                break
            end if
        end while
        while n(Di+1) > k do
            n = n(Di)
            m = n(Di+1) - 1
            X = - $\frac{n}{n+1}(x - \bar{x})^2 + \frac{m+1}{m}(x - \bar{y}')^2$ 
            if X > 0 then
                Move x from Di+1 to Di
                flag ← 1
            else
                break
            end if
        end while
    end for
end while

```

ズム MIL を実行し、情報損失を比較した結果について説明する。実験は以下の手順で行う。

- (1) 様々な確率分布に従う乱数のデータセットを生成する。
- (2) MDAV, MDAV+MIL, VMDAV, VMDAV+MIL を実行してクラスタを生成する。
- (3) 実行結果の情報損失を比較する。
- (4) MIL のデータ移動の際の比較回数を数えあげ、実行時間を調べる。

4.1 テストデータセットの生成

アルゴリズムの実行のため、表 1 の確率密度関数に従う乱数のデータセットを用意した。ただし、 $N(\mu, \sigma)$ は、平均 μ 、分散 σ^2 の正規分布を表す。確率密度関数は、実データを表す最も一般的な分布である正規分布のほかに、複数の偏りを持つ分布を用意し、様々な事象を想定したデータセット生成を行う。

まず、正規乱数のサンプルデータセットを生成する方法について述べる。一様分布データ d 個の平均値 \bar{X} の確率分布は、中心極限定理により近似的に正規分布に従うので、サンプルデータは、

$$z(x) = \frac{\bar{X} - \frac{1}{2}}{\sqrt{\frac{1}{12d}}}$$

で算出できる [9]。平均が μ 、分散が σ^2 である確率分布 $f(x)$ の場合、サンプルデータは $z_i(x) = \sigma z(x) + \mu$ となる。

表 1 データセットの確率分布

Table 1 The probability distributions of the data sets.

データセット	確率密度関数	データ総数
DS0	$N(0, 1)$	100
DS1	$\frac{1}{2}N(5, 1) + \frac{1}{2}N(10, 1)$	200
DS2	$\frac{1}{2}N(5, 1) + \frac{1}{2}N(8, 1)$	200
DS3	$\frac{1}{2}N(5, 1) + \frac{1}{2}N(10, 2)$	200
DS4	$\frac{1}{2}N(10, 3) + \frac{1}{2}N(20, 2)$	200
DS5	$\frac{1}{3}N(0, 1) + \frac{1}{3}N(5, 2) + \frac{1}{3}N(12, 3)$	300
DS6	$\frac{1}{3}N(5, 1.5) + \frac{1}{3}N(10, 1) + \frac{1}{3}N(15, 1.5)$	300
DS7	$\frac{1}{3}N(5, 3) + \frac{1}{3}N(15, 2) + \frac{1}{3}N(20, 1)$	300
DS8	$\frac{1}{3}N(5, 3) + \frac{1}{3}N(12, 1.5) + \frac{1}{3}N(20, 2)$	300
DS9	$\frac{1}{3}N(5, 2) + \frac{1}{3}N(10, 1.5) + \frac{1}{3}N(18, 3)$	300
DS10	$\frac{1}{3}N(0, 1) + \frac{1}{3}N(5, 1) + \frac{1}{3}N(10, 1)$	300
DS11	$\frac{1}{3}N(0, 1) + \frac{1}{3}N(3, 1) + \frac{1}{3}N(6, 1)$	300

このようにして算出したデータの集合を正規乱数のサンプルデータセットとする。本論文では、 $d = 6$ として正規乱数列を生成する。

次に、異なる正規乱数を重ね合わせた確率分布に従うサンプルデータセットの生成について述べる。 w 個の確率分布を重ね合わせる場合、 $f_1(x)$ に従うサンプルデータセットを x_1, \dots, x_{n_1} 、 $f_2(x)$ に従うサンプルデータセットを y_1, \dots, y_{n_2} というように、各確率分布に対して生成したサンプルデータセットを合わせたデータセット $x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2}, \dots$ は、確率分布 $f(x) = \frac{1}{\sum_{i=1}^w n_i} (n_1 f_1(x) + \dots + n_w f_w(x))$ に従うサンプルデータセットである [10]。

4.2 アルゴリズムの実行

前節で示したデータセットを用いて、既存アルゴリズム MDAV, VMDAV と提案アルゴリズムを実行し、情報損失を比較する。アルゴリズムは、データセットおよび k 値を与えて実行し、データ総数を N とすると、 k 値は $2 \sim \frac{N}{2}$ とする。提案アルゴリズムの実行には、MDAV および VMDAV により生成されたクラスタを使用する。また、VMDAV の実行では、アルゴリズムのパラメータ γ は 1.0 とする。

4.3 MIL による情報損失減少可能性の検証

まず、標準正規分布のデータセット DS0 にアルゴリズムを実行した結果を述べる。次に、より複雑な分布の代表例として DS1, DS9 についてアルゴリズムを実行した結果を述べるが、他のデータセットに対しても類似の結果であった。最後に、結果に対する考察を述べる。

標準正規分布である DS0 について、情報損失の結果を図 1, 図 2 に示す。また、 $k = 2 \sim 15$ の場合の情報損失を表 2 に示す。MDAV について、 $k = 2 \sim 50$ のうち 69.4% が MIL により改善され、その際の情報損失の減少率は最大 7.6% であった。また、VMDAV については 91.8% が改善さ

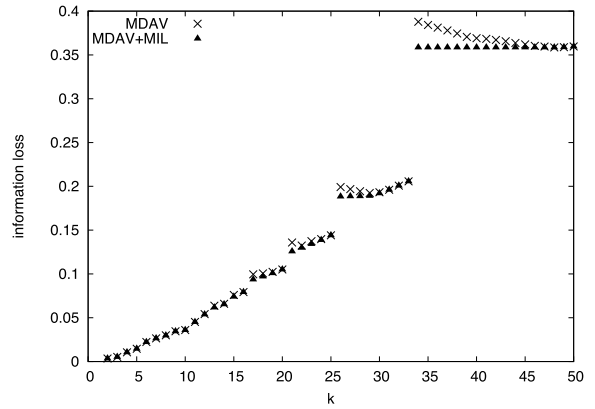


図 1 DS0 の情報損失の比較 (MDAV)

Fig. 1 Information loss for DS0 (comparison with MDAV).

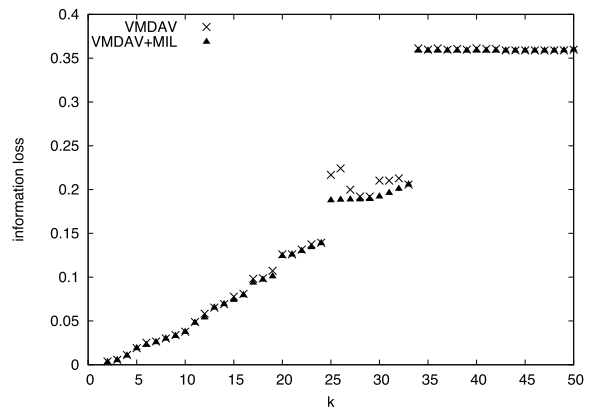


図 2 DS0 の情報損失の比較 (VMDAV)

Fig. 2 Information loss for DS0 (comparison with VMDAV).

表 2 DS0 の情報損失の値

Table 2 Information loss for DS0.

k	MDAV	MDAV+MIL	VMDAV	VMDAV+MIL
2	0.003446	0.003446	0.003765	0.002946
3	0.005258	0.005258	0.005632	0.005282
4	0.010552	0.010552	0.011109	0.010479
5	0.014560	0.014560	0.018915	0.018828
6	0.022300	0.022263	0.025131	0.022560
7	0.026470	0.026470	0.026578	0.025728
8	0.029978	0.029778	0.030077	0.029778
9	0.034523	0.034490	0.034029	0.032834
10	0.035952	0.035952	0.037727	0.037490
11	0.045240	0.045126	0.048749	0.047834
12	0.054165	0.053752	0.058318	0.053937
13	0.063774	0.061702	0.065301	0.065207
14	0.065962	0.065343	0.069149	0.069149
15	0.076020	0.074068	0.077760	0.074068

れ、情報損失の減少率は最大 21.7% であった。

2つの分布の偏りの密度が同一である DS1 について述べる。確率分布を図 3 に、情報損失の結果を図 4, 図 5 に示す。また、 $k = 2 \sim 15$ の場合の情報損失を表 3 に示す。MDAV について、 $k = 2 \sim 100$ のうち 89.9% が MIL により

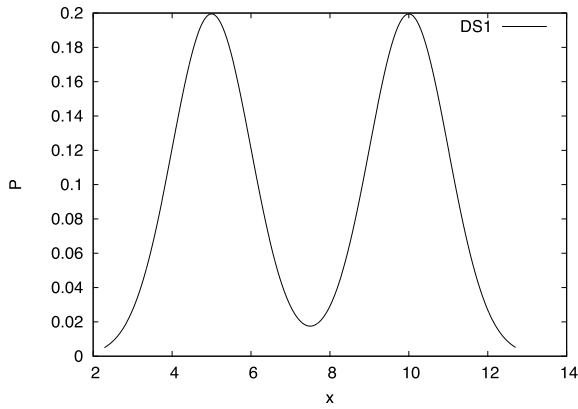


図 3 DS1 の確率分布

Fig. 3 Probability distribution of DS1.

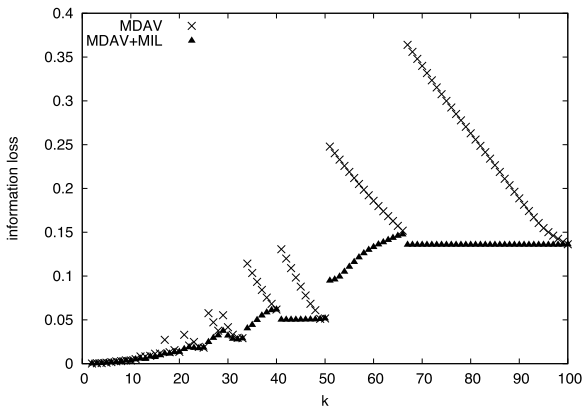


図 4 DS1 の情報損失の比較 (MDAV)

Fig. 4 Information loss for DS1 (comparison with MDAV).

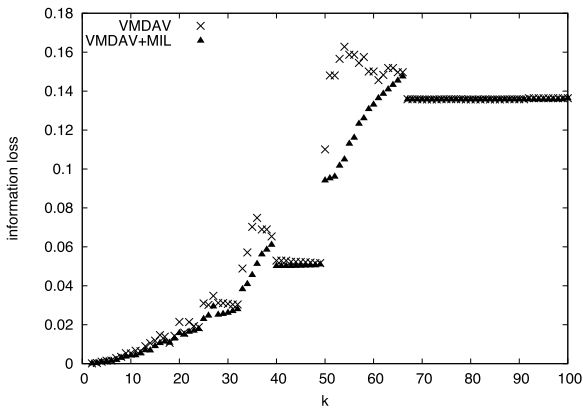


図 5 DS1 の情報損失の比較 (VMDAV)

Fig. 5 Information loss for DS1 (comparison with VMDAV).

改善され、その際の情報損失の減少率は最大 64.7%であった。また、VMDAV については 86.9%が改善され、情報損失の減少率は最大 35.8%であった。

より複雑な確率分布に従うデータセットの一例として DS9 の実験結果を述べる。確率分布を図 6 に、情報損失の結果を図 7、図 8 に示す。また、 $k = 2 \sim 15$ の場合の情報損失を表 4 に示す。MDAV について、 $k = 2 \sim 150$ のうち 78.5%が MIL により改善され、その際の情報損失の減少率は

表 3 DS1 の情報損失の値

Table 3 Information loss for DS1.

k	MDAV	MDAV+MIL	VMDAV	VMDAV+MIL
2	0.000214	0.000214	0.000143	0.000139
3	0.000489	0.000405	0.000350	0.000324
4	0.000651	0.000651	0.001027	0.000892
5	0.001114	0.001114	0.001530	0.001244
6	0.001777	0.001639	0.001598	0.001452
7	0.002643	0.001780	0.002891	0.001921
8	0.002895	0.002895	0.003499	0.002982
9	0.003749	0.003405	0.005321	0.003599
10	0.003635	0.003635	0.005246	0.004195
11	0.005227	0.004887	0.006654	0.004279
12	0.008307	0.005765	0.006414	0.005243
13	0.008298	0.005407	0.009004	0.006870
14	0.008943	0.007969	0.010573	0.006791
15	0.011400	0.007469	0.011821	0.008994

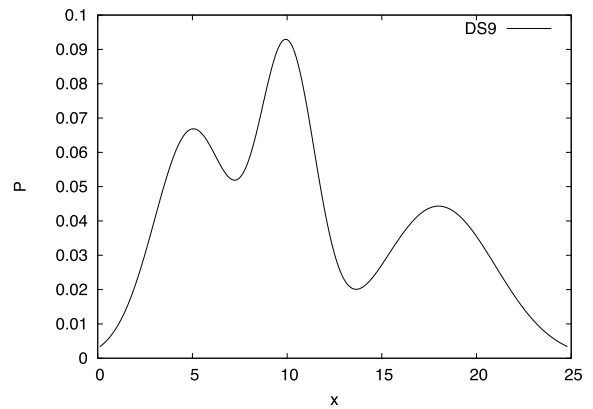


図 6 DS9 の確率分布

Fig. 6 Probability distribution of DS9.

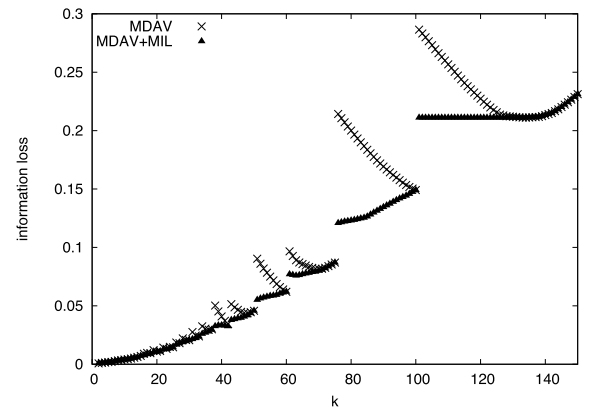


図 7 DS9 の情報損失の比較 (MDAV)

Fig. 7 Information loss for DS9 (comparison with MDAV).

最大 43.6%であった。また、VMDAV については 98.7%が改善され、情報損失の減少率は最大 24.0%であった。

DS0~DS11 に関して、MDAV について、 $k = 2 \sim \frac{N}{2}$ のうち 49.7%~89.9%が MIL により改善され、その際の情報損失の減少率は最大 67.3%であった。また、VMDAV につ

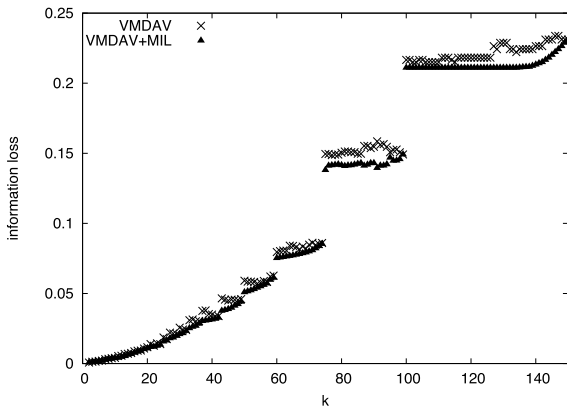


図 8 DS9 の情報損失の比較 (VMDAV)

Fig. 8 Information loss for DS9 (comparison with VMDAV).

表 4 DS9 の情報損失の値

Table 4 Information loss for DS9.

k	MDAV	MDAV + MIL	VMDAV	VMDAV + MIL
2	0.000682	0.000682	0.000685	0.000681
3	0.001096	0.001096	0.001284	0.001128
4	0.001402	0.001402	0.001423	0.001370
5	0.001848	0.001848	0.001868	0.001717
6	0.002002	0.002002	0.002134	0.002053
7	0.002668	0.002634	0.003125	0.002620
8	0.003260	0.003175	0.003406	0.003220
9	0.003535	0.003461	0.003809	0.003506
10	0.004043	0.004043	0.04060	0.003910
11	0.004306	0.004241	0.005335	0.004055
12	0.004870	0.004870	0.005596	0.004704
13	0.005554	0.005515	0.006799	0.005304
14	0.006240	0.005791	0.007001	0.006234
15	0.007143	0.007143	0.007538	0.006761

いては 70.7%~99%が改善され、情報損失の減少率は最大 51.7%であった。

これらの実験結果についての考察を述べる。k が小さい場合や、データ総数を k で割った余りが小さい場合には、MDAV および VMDAV の情報損失は、MIL によりほとんど改善されなかった。MIL はデータ数が k より大きいクラスタからのデータ移動を行うアルゴリズムであり、この場合にはデータ移動の自由度が低いと考えられる。とくに、データ総数が k の倍数のときは、MDAV によってデータ数 k のクラスタに分割されるため、データ移動による情報損失改善の余地はない。一方、データ総数を k で割った余りが大きいときには、MIL により情報損失が大きく改善される場合が多く見られた。

4.4 MIL の実行時間

データ総数が異なるデータセットに対する、アルゴリズムの実行時間について調べる。実行時間は、クラスタ間をデータが移動する際の判定回数を用いて評価を行う。実験

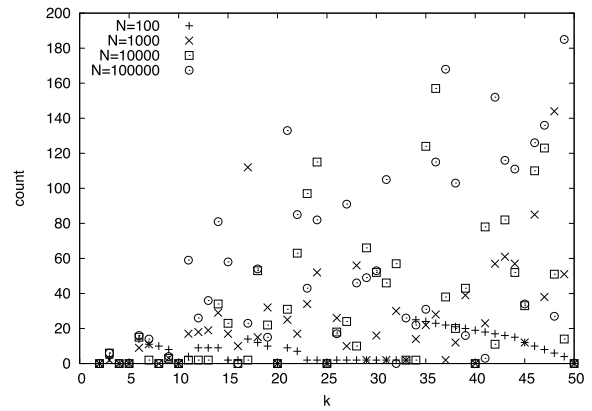


図 9 Seed0 の判定回数

Fig. 9 The judgement count.

は以下の手順で行う。

- (1) 4.2 節のようにして標準正規分布に従うデータ数 N が 100, 1,000, 10,000, 100,000 のデータセットを生成する。データセットは、乱数のシードを変えて、3 セットずつ用意する。
- (2) (1) で生成したデータセットについて MDAV を実行しクラスタを生成する。生成したクラスタについて MIL を実行し、データ移動の判定回数を $k = 2 \sim 50$ についてカウントする。

Seed0 のデータセットに対する判定回数をプロットした結果を図 9 に示す。3 つのデータセットに対する判定回数は、 $N = 100$ のとき最大 37 回、 $N = 1,000$ のとき最大 144 回、 $N = 10,000$ のとき最大 214 回、 $N = 100,000$ のとき最大 185 回であった。

MDAV によるクラスタリングの結果、クラスタのデータ数は $k, \dots, k, l, k, \dots, k$ ($k \leq l < 2k$) のようになる。MIL によりデータ数 l のクラスタの $(l - k)$ 個以下のデータがその付近の $(l - k)$ 個以下のクラスタに移動する。また、判定が必要なのはそれらのクラスタのみである。したがって、判定回数は k のみに依存しデータ総数には依存せず、データ総数が大きくなっても MIL は高速に実行可能である。

5. まとめ

本論文では、k-匿名性を持つ匿名化情報に対して、クラスタ間でデータの移動を行うことにより情報損失を極小にするアルゴリズム MIL を開発し、その実装および実験、評価を行った。その結果、データ総数を k で割った余りが大きいとき、多くの場合、MDAV, VMDAV による情報損失を MIL によって改善可能であった。さらに、MIL の実行時間について評価した結果、データ数が大きくなってもデータ数によらず高速に実行可能であることが分かった。

参考文献

[1] Sweeney, L.: k-anonymity: A model for protecting privacy, *International Journal of Uncertainty, Fuzziness*

and *Knowledge-Based Systems*, Vol.10, No.5, pp.557-570 (2002).

- [2] 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, Springer (2005).
- [3] Solanas, A., Martinez-Balleste, A. and Domingo-Ferrer, J.: V-MDAV: A multivariate microaggregation with variable group size, *17th COMPSTAT Symposium of the IASC*, Rome (2006).
- [4] Edwards, A.W.F. and Cavalli-Sforza, L.L.: A method for cluster analysis, *Biometrics*, Vol.21, pp.362-375 (1965).
- [5] Gordon, A.D. and Henderson, J.T.: An algorithm for Euclidean sum of squares classification, *Biometrics*, Vol.33, pp.355-362 (1977).
- [6] Hansen, P., Jaumard, B. and Mladenovic, N.: Minimum sum of squares clustering in a low dimensional space, *Journal of Classification*, Vol.15, pp.37-55 (1998).
- [7] MacQueen, J.: Some methods for classification and analysis of multivariate observations, *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, pp.281-297 (1967).
- [8] Ward, J.H.: Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, Vol.58, pp.236-244 (1963).
- [9] 脇本和昌: 乱数の知識, pp.74-82, 森北出版 (1970).
- [10] 平岡和幸, 堀 玄: プログラミングのための確率統計, pp.113-132, オーム社 (2009).



秋山 寛子 (正会員)

国立長野工業高等専門学校電子情報工学科助教。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。電子情報通信学会会員。



和田 昌昭

大阪大学大学院情報科学研究科情報基礎数学専攻教授。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年より現職。バイオイメージング学会会員。



中山 雅哉 (正会員)

東京大学情報基盤センター准教授。1989年東京大学大学院工学系研究科情報工学専攻博士課程修了(工学博士)。広域分散処理技術に関する研究に従事。IEEE, 電子情報通信学会各会員。



加藤 朗 (正会員)

慶應義塾大学大学院メディアデザイン研究科教授。東京工業大学工学部理工学研究科情報工学専攻後期博士課程満期退学。1990年慶應義塾大学湘南藤沢キャンパスのキャンパスネットワーク構築準備に携わり, 同キャンパスにて環境情報学部助手。1990年東京大学大型計算機センター(現, 情報基盤センター)助手, 同助教授(現, 准教授)を経て, 2008年より現職。博士(政策・メディア)。



砂原 秀樹 (正会員)

慶應義塾大学大学院メディアデザイン研究科教授。慶應義塾大学先端研究センターサイバーセキュリティ研究センター所長。1988年慶應義塾大学理工学部博士課程修了。電気通信大学情報工学科助手, 1994年奈良先端科学技術大学院大学情報科学センター助教授を経て, 2001年から教授。2005年情報科学研究科教授。2008年4月より現職。村井純(慶應義塾大学環境情報学部教授)らとともに, 1984年からJUNET, 1988年からWIDEプロジェクトを通じて, 日本におけるインターネットの構築とその研究に従事。2005年より江崎浩(東京大学教授)とともにインターネットを通じて環境情報を共有するLive E! Projectを開始。一方, 2008年よりIT Keys, 2013年よりenPiT Security/SecCapプログラムを実施している。このほか, 自動車や様々なセンサをインターネットに接続して新たな情報通信基盤を構築するプロジェクト, パーソナル情報を安心・安全に活用するためのフレームワーク「情報銀行」に関するプロジェクトを進行中。