情報処理学会論文誌　データベース　Vol.6　No.5　38–48 (Dec. 2013)

センシティブ属性間の関係多様化によるプライバシ保護手法

髙橋 翼[1,a)]　側高 幸治[1]　竹之内 隆夫[1]　森 拓也[1]

受付日 2013年6月21日, 採録日 2013年10月7日

概要：本稿では，2 つのセンシティブ属性を持つレコード群に対するプライバシ保護の問題を扱う．人々の生活や行動を記録したパーソナル情報には一データ主体のレコードに複数のセンシティブ属性が記録されたものが存在する．これらのパーソナル情報からはセンシティブ属性間の関係を分析することができる．一方で，特定のデータ主体に関するあるセンシティブ属性に関する知識から他のセンシティブ属性値が特定されるプライバシ侵害が生じうる．このようなプライバシ侵害を防ぐためにはセンシティブ属性間の関係の多様化が必要だが，関係の曖昧化が生じてデータ分析精度を劣化させる可能性がある．本稿では，2 つのセンシティブ属性を持つデータセットに対して，関係の曖昧化を抑制しつつ効率的に関係多様化を実現する方式を提案する．評価実験では，提案手法がセンシティブ属性間の関係の過度な曖昧性を抑止しながら関係多様化を実現でき，高い効率性を有することを示す．

キーワード：データベースプライバシ，データ匿名化，関係多様化

Privacy Preserving Publishing of Sensitive Attributes in Binary Relations

Tsubasa Takahashi[1,a)]　Koji Sobataka[1]　Takao Takenouchi[1]　Takuya Mori[1]

Received: June 21, 2013, Accepted: October 7, 2013

Abstract: This paper proposes a data anonymization method for records having sensitive attributes in binary relations. If a record is identified by exploiting knowledge about an sensitive attribute, the values of the other sensitive attributes are revealed. In order to prevent such revealing of the sensitive attributes, we propose relation diversity that is the metric of mutual relationship diversity among sensitive attributes in binary relations. However, ensuring relation diversity makes the relations among sensitive attributes obfuscated. Further this paper proposes a data anonymization which ensures relation diversity with small obfuscation in an efficient way. Experimental evaluations show the effectiveness of our methods.

Keywords: database privacy, data anonymization, relation diversity

1. はじめに

診療履歴やサイト訪問履歴といったパーソナル情報が，サービスを受けるたびに蓄積されている．近年，ビッグデータ活用のニーズが高まり，これらの蓄積されたパーソナル情報を第三者のサービスや事業に活用する二次活用の期待が高まっている．パーソナル情報の複数の属性の関係から，属性間の相関や変化を観察することができる．たとえば，**表 1** (a) は傷病と薬剤を記録したテーブルであり，傷病と薬剤との相関が得られる．表 1 (b) と表 1 (c) の時系列データからは 4 月と 5 月の傷病の時間変化が得られる．しかしながら，傷病や薬剤のようにデータ主体に関する機微な情報（センシティブ属性）は，第三者に知られたくない情報であるため，二次活用の際にはデータ主体のプライバシへの配慮が必要となる．

データ主体のプライバシを保護する技術として，データセットに所望の匿名性の指標を充足させるデータ匿名化が知られている．データ匿名化手法の 1 つである k-匿名化 [1] は，個人を特定しうる属性（準識別子）を加工して同一の準識別子の組を持つレコードが k 個以上出現することを保

[1] 日本電気株式会社
NEC Corporation, Kawasaki, Kanagawa 211–8666, Japan
[a)] t-takahashi@nk.jp.nec.com

表 1　パーソナルデータ
Table 1　Personal data.

(a)

ID	年代	傷病	薬剤
1	40代	A	X
2	40代	B	Y
3	40代	A	Y
4	40代	B	X
5	40代	C	Z
6	40代	C	X

(b)

ID	診療月	傷病
1	4	A
2	4	B
3	4	A
4	4	B
5	4	C
6	4	C

(c)

ID	診療月	傷病
1	5	C
2	5	D
3	5	D
4	5	C
5	5	Z
6	5	X

表 2　関係多様化データ
Table 2　Relational diverse data.

(a)

年代	傷病	薬剤
40代	{A,B}	{X,Y}
40代	{A,B}	{X,Y}
40代	{A,C}	{X,Y}
40代	{B,C}	{X,Z}
40代	{B,C}	{X,Z}
40代	{A,C}	{X,Y}

(b)

GID	診療月	傷病
G1	4	A
G1	4	B
G2	4	A
G3	4	B
G3	4	C
G2	4	C

(c)

GID	診療月	傷病
G1	5	C
G1	5	D
G2	5	D
G3	5	C
G3	5	Z
G2	5	X

証する技術であり，k-匿名化によって準識別子を知識とした個人特定が困難になる．また，準識別子の組から特定しうるセンシティブ属性値を複数種類であることを保証する処理を多様化と呼び，多様性の指標として ℓ-多様化 [2] が提案されている．

前述のように複数のセンシティブ属性を含むパーソナル情報では，特定のデータ主体に関するあるセンシティブ属性に関する知識から他のセンシティブ属性値が特定されるプライバシ侵害が生じうる．このプライバシ侵害を防ぐためには，あるセンシティブ属性から他のセンシティブ属性が一意に対応付かないように，あるセンシティブ属性から他のセンシティブ属性への対応関係が多様になることを保証する必要がある．本稿では，センシティブ属性間の対応付けの多様合いを関係多様性と呼び，所定の関係多様性を満たすための操作を関係多様化と呼ぶ．

たとえば，表 1 (a) は**表 2** (a) のように，表 1 (b) と表 1 (c) の時系列データは表 2 (b) と表 2 (c) のように加工されることでセンシティブ属性間の関係が多様化される．ここで，表 2 (a) は，センシティブ属性値を複数の値を含む集合へと加工することで，センシティブ属性間の関係を多様化している．また，表 2 (b) と表 2 (c) では，複数のセンシティブ属性値が含まれるようにレコード群をグループ化し，センシティブ属性間の対応関係をグループ間の対応関係へと曖昧化することで関係多様化を実現している．

ただし，関係多様化はセンシティブ属性間の対応関係に曖昧性を生じさせるため，分析精度を劣化させる可能性がある．本稿では，2 つのセンシティブ属性を持つパーソナル情報を対象として，関係多様性を関係の曖昧化の度合いを抑制しながら実現するデータ匿名化に取り組む．まず，センシティブ属性間の関係多様性である (ℓ_1, ℓ_2)-関係多様性を提案する．続いて，(ℓ_1, ℓ_2)-関係多様化の際に生じる関係の曖昧化について議論し，関係の曖昧化を抑止した (ℓ_1, ℓ_2)-関係多様化手法を提案する．さらに，関係の曖昧化を効果的に抑制しつつ効率的に (ℓ_1, ℓ_2)-関係多様化を実現する手法の提案も行う．評価実験では，提案手法が関係の曖昧性を抑止しつつ効率良く (ℓ_1, ℓ_2)-関係多様化を実現できることを示す．

本稿の以降の構成は以下のとおりである．2 章では，本稿の論述に必要な基本的事項の導入を行う．3 章では，センシティブ属性の関係多様性の指標である (ℓ_1, ℓ_2)-関係多様性を提案する．4 章では，関係の抽象化を抑止した (ℓ_1, ℓ_2)-関係多様化手法を提案する．5 章では，いくつかのヒューリスティクスを用いた効率的な (ℓ_1, ℓ_2)-関係多様化手法を提案する．6 章では，提案手法の有効性について評価実験を通して論じる．7 章において関連研究を紹介し，最後に 8 章で，本稿の結論を述べる．

2. 準備

2.1 対象とするデータ

本稿で対象とするデータは，テーブル型のデータベースである．テーブル T をなすタプル t は，準識別子 $QI_1, \ldots, QI_m$ と，2 つのセンシティブ属性 S_1 と S_2 を含む．タプル t の属性 S_i の値を $t.S_i$ とする．

表 1 (b) と表 1 (c) のように，同一のデータ主体に対して 2 つ以上のレコードが存在する場合，それらを結合したテーブルも対象の 1 つである．

2.2 攻撃モデル

攻撃者として，特定の個人の S_1 と S_2 のいずれかの属性値と，m 準識別子の値 $qi_1, \ldots, qi_m$ $(qi_j \in QI_j\,(j = 1, \ldots, m))$ を知識として有し，それらを用いて未知のセンシティブ属性の値を特定しようとする者を想定する．

ここで s_k と $qi_1, \ldots, qi_m$ を持つタプル集合を $T(s_k, qi_1, \ldots, qi_m)$ とする．また，s_1 と $qi_1, \ldots, qi_m$ を持つタプル集合の S_2 の値の集合を $S_2(s_1, qi_1, \ldots, qi_m)$ とする．

$$T(s_k, qi_1, \ldots, qi_m) = \{t | \forall t \in T \ s.t. \ t.S_k = s_k \wedge qi_j(t) = qi_j \wedge j = 1, \ldots, m\} \quad (1)$$

$$S_1(s_2, qi_1, \ldots, qi_m) = \{s_1(t) | \forall t \in T(s_2, qi_1, \ldots, qi_m)\} \quad (2)$$

$$S_2(s_1, qi_1, \ldots, qi_m) = \{s_2(t) | \forall t \in T(s_1, qi_1, \ldots, qi_m)\} \quad (3)$$

攻撃として，特定個人の未知のセンシティブ属性値の候補を σ（≥ 1）種類以下に絞り込むことを想定する．

上述の攻撃者からセンシティブ属性値の特定を防ぐため

には，同一の準識別子の組を持つレコード群に対して S_1，S_2 それぞれをキーとして検索可能な他方の値を一定の種類以上であることを保証する必要がある．

なお，本稿で想定する攻撃モデルは従来の ℓ-多様性が想定する確率的な攻撃モデルとは異なる．従来の ℓ-多様性はセンシティブ属性値をある確率以上で推定するといった確率的な攻撃を想定しているが，本稿では特定個人のセンシティブ属性値をある一定の種類数以下に絞り込むことを想定している．

3. 関係多様性と関係曖昧性指標

2.2 節の攻撃モデルによるプライバシ侵害を防ぐために，テーブルが満たすべき関係多様性の指標（3.1 節）と，関係多様化による関係の曖昧性指標（3.2 節）を導入する．

3.1 提案指標：(ℓ_1, ℓ_2)-関係多様性

二項関係にある 2 つのセンシティブ属性を持つテーブルに対する，一方のセンシティブ属性値から特定可能な他方のセンシティブ属性値の種類数を表す関係の多様性指標 (ℓ_1, ℓ_2)-関係多様性を定義する．

定義 1　（(ℓ_1, ℓ_2)-関係多様性）$\forall t \in T$, $qi_i = qi_i(t)$, $s_j = s_j(t)$ に対して，$|S_1(s_2, qi_1, \ldots, qi_m)| \geq \ell_1$ と $|S_2(s_1, qi_1, \ldots, qi_m)| \geq \ell_2$ が成り立つとき，テーブル T は (ℓ_1, ℓ_2)-関係多様性を満たす．

(ℓ_1, ℓ_2)-関係多様性を満たすテーブルの例を表 2 (a) と表 2 (b)，表 2 (c) に示す．ここで，同一の $(qi_1, \ldots, qi_m)$ を持つタプルの集合をクラスと呼ぶ．表 2 (b)，表 2 (c) のような形式で関係多様化された場合は，同一のグループ識別子 GID を付与されているタプルの集合をクラスと呼ぶ．クラス c 中のすべてのタプルが $|S_1(s_2, qi_1, \ldots, qi_m)| \geq \ell_1$ と $|S_2(s_1, qi_1, \ldots, qi_m)| \geq \ell_2$ を満たすとき，クラス c は (ℓ_1, ℓ_2)-関係多様性を満たすという．また，(ℓ_1, ℓ_2)-関係多様性を満たすクラスを (ℓ_1, ℓ_2)-関係多様なクラスと呼ぶ．

なお，以降議論を簡単化するために，クラス内のタプルの準識別子は同一の値に加工されるものとし，本稿では，準識別子の情報損失や歪曲度等については議論の対象としない．

3.2 関係の曖昧性指標

(ℓ_1, ℓ_2)-関係多様性を保証する (ℓ_1, ℓ_2)-関係多様化を行うと，S_1 の値はサイズ ℓ_1 以上の集合に，S_2 の値はサイズ ℓ_2 以上の集合となる．よって，(S_1, S_2) の二項関係は，集合間の二項関係へと曖昧化される．

たとえば，(3, 2)-関係多様性を保証した場合，$(\{a, b, c\}, \{x, y\})$ なる二項関係を持つ (3, 2)-関係多様化されたタプル集合（クラス）が存在することを考える．このクラスをなすタプルのオリジナルの関係がそれぞれ (a, x)，(b, y)，(c, x) とする．このとき，$(\{a, b, c\}, \{x, y\})$ には，オリジナルの関係には存在しない関係（(a, y)，(b, x)，(c, y)）が含まれていることが分かる．関係多様化によって混入するオリジナルの関係には存在しない関係をノイズ関係（Noisy Relation）と呼び，以降，単にノイズと呼ぶ．一方，クラスをなすタプルの関係が (a, x)，(a, y)，(b, x)，(b, y)，(c, x)，(c, y) の場合には，クラスの二項関係（$\{a, b, c\}, \{x, y\}$）は，オリジナルのすべての関係で成り立っている．ノイズが混入すると，様々なデータ分析の精度に影響を与える可能性があるため，後者の例のようにノイズが混入しないことが望ましい．

ここで，クラス c に含まれる S_k の集合を $S_k(c)$ とする（$k \in \{1, 2\}$）．

$$S_k(c) = \{t.S_k | \forall t \in c\} \tag{4}$$

また，クラス c のメンバであるタプルのセンシティブ属性間のオリジナルの関係（s_1，s_2）の集合を $R(c)$ とする．

$$R(c) = \{(s_1(t), s_2(t)) | \forall t \in c\} \tag{5}$$

関係多様化によって，$R(c)$ は $S_1(c)$ と $S_2(c)$ の値の組合せからなる関係の集合へと曖昧化される．曖昧化された関係の集合を $R^*(c)$ とする．$R^*(c)$ の関係数は $R^*(c) = |S_1(c)||S_2(c)|$ で与えられる．

以上より，クラス中のオリジナルの関係とノイズとの比を表す関係ノイズ比 RNR を以下のように定義する．

定義 2　（関係ノイズ比）

$$RNR(c) = \frac{|S_1(c)||S_2(c)|}{|R(c)|} \tag{6}$$

$RNR(c)$ は 1 以上の値をとり，最小値（1）のとき，クラスにノイズが混入していないことを表す．ノイズの混入のないクラスをノイズレスクラスと呼ぶ．

本稿の関係ノイズ比は，クラス単位の局所的な視点に基づくものであり，テーブル全体の関係ノイズ比を全域的な視点での評価したものではない．また，本稿で定義した関係の曖昧性指標はノイズの頻度や偏りをとらえられておらず，それらによる関係の曖昧化を小さく見積もってしまう場合がある．本稿では，評価指標を簡略化するためにノイズの有無による指標を用いた．ノイズの頻度や偏りを考慮した評価指標の確立は今後の課題とする．

4. (ℓ_1, ℓ_2)-関係多様化

関係の曖昧化を抑止した関係多様化テーブルの生成について考える．最適な関係多様化は，関係多様化後の各タプルの関係ノイズ比の総和が最少となる関係多様化である．ただし，最適な k-匿名化が NP 困難な問題 [3] であることから，k-匿名化を前提とする関係多様化についても最適解の導出は NP 困難と考えられる．

そこで本稿では，ヒューリスティックな手法によってで

きるだけ関係ノイズ比が小さいクラスを生成することを考える．ここでは，クラスタリングによって関係多様化を実現する手法を提案する．4.1 節ではクラスタリングによる関係多様化手法を述べ，4.2 節では関係ノイズ比を考慮した類似度指標を述べる．

4.1　クラスタリングによる関係多様化

まず，関係の曖昧性については考慮せず，タプル群のクラスタリングによって関係多様化を実現する手法を提案する．このクラスタリング手法では，(ℓ_1, ℓ_2)-関係多様性に近づくように 2 つのクラスを順次併合するという方針をとる．

クラスタリング手法として，凝集型の階層的クラスタリング [4] を用いる．関係多様化に用いる凝集型の階層的クラスタリングは，以下のステップのアルゴリズムを採用する．

- ステップ 1：各関係から 1 つの関係だけを含むクラス c を生成．
- ステップ 2：クラス間の類似度 $sim(c_1, c_2)$ を計算し，類似度が 0 より大きいペアの中で最も類似度の高いペアを併合．
- ステップ 3：併合したクラスがノイズレスクラスになったらクラスタリングの対象から除く．
- ステップ 4：ステップ 2 で併合が行われればステップ 2〜3 を繰り返し，併合が行われなければ終了．

クラス間の類似度 $sim(c_1, c_2)$ は，2 つのクラスを併合した場合における (ℓ_1, ℓ_2)-関係多様性の充足性（充足の度合い）によって評価する．

ここで，クラスのセンシティブ属性 S_k の多様性は式 (7) で表す．

$$div_k(c) = |\{t.S_k | t \in c\}| \tag{7}$$

2 つのクラス c_1 と c_2 を併合した際のセンシティブ属性 S_i の多様性は式 (8) で導出できる．

$$div_k(c_1, c_2) = min(\ell_i, |\{t.S_k | t \in c_1 \cup c_2\}|) \tag{8}$$

さらに，c_1 と c_2 を併合した際の多様性の変化量は式 (9) で表すことができる．

$$\triangle div_k(c_1, c_2) = div_k(c_1, c_2) - max(div_k(c_1), div_k(c_2)) \tag{9}$$

$\triangle div_k(c_1, c_2)$ が正のとき，クラス c_1 と c_2 を併合することで，(ℓ_1, ℓ_2)-関係多様なクラスに近づくことができる．(ℓ_1, ℓ_2)-関係多様性の充足性を式 (10) で表す．

$$rdiv(c_1, c_2) = \frac{div_1(c_1, c_2) + div_2(c_1, c_2)}{\ell_1 + \ell_2} \tag{10}$$

$rd = rdiv(c_1, c_2)$ は $2/(\ell_1 + \ell_2) \leq rd \leq 1$ の値をとり，1 のとき (ℓ_1, ℓ_2)-関係多様性を満たすことを表す．よって，$rdiv$ 値が高いクラスペアどうしを併合すると，(ℓ_1, ℓ_2)-関係多様性の充足に大きく近づく．

これまでの議論より，(ℓ_1, ℓ_2)-関係多様性の充足性の高いクラスペアを発見するためのクラス間の類似度指標として，以下の式 (11) の DG（Diversity Gain）を導入し，凝集型の階層的クラスタリングで関係多様化を行う．

$$DG(c_1, c_2) = \begin{cases} rdiv(c_1, c_2) & (\triangle div_1(c_1, c_2) > 0) \\ rdiv(c_1, c_2) & (\triangle div_2(c_1, c_2) > 0) \\ 0 & (otherwise) \end{cases} \tag{11}$$

4.2　関係ノイズ比を考慮した関係多様化

関係ノイズ比を考慮したクラス間の評価指標を考える．まず，関係ノイズ比 RNR を考慮して，関係の損失を表す評価値 RL（Relation Loss）を導入する．

$$RL(c_1, c_2) = exp(RNR(c_1 \cup c_2) - 1) \tag{12}$$

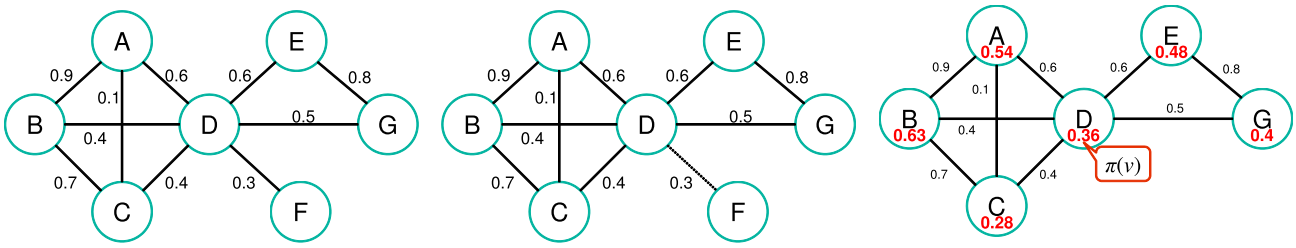
DG と RL を用いて，関係多様性の充足率と，関係ノイズ比の両方を加味した評価値 $DGRL$ を式 (13) のように定義する．

$$DGRL(c_1, c_2) = \frac{DG(c_1, c_2)}{RL(c_1, c_2)} \tag{13}$$

上述の $DGRL$ をクラス間類似度 $sim(c_1, c_2)$ として凝集型の階層的クラスタリングを用いることで，関係ノイズ比が小さい (ℓ_1, ℓ_2)-関係多様なクラスが次々に生成される．

例 1　(a, x)，(b, y)，(a, y)，(b, x) をセンシティブ属性として持つタプルを対象にして，(2, 2)-関係多様化させる場合を考える．(a, x)，(b, y) の併合は，DG 値が 2 の併合であり，それぞれ (2, 2)-関係多様性を満たす．しかし，関係ノイズ比 RNR は 2 であり，各クラスにはノイズ関係が混入してしまう．さらにこの併合は RL 値は $exp(2-1) > 2$ であり，$DGRL$ 値は $2/e$ である．一方で，(a, x) と (a, y) の併合は，ノイズ関係を混入せずに (1, 2)-関係多様性を得られる．このとき，DG 値はともに 2，RL 値はともに 1 であり，$DGRL$ 値は 1 である．よって，$DGRL$ 値の高い後者の組合せを採用し，同様に (b, x) と (b, y) を併合する．続いて，併合された 2 つのクラスをさらに併合すると，ノイズ関係のない (2, 2)-関係多様化を実現できる．

凝集型の階層的クラスタリングは各ステップで最良の評価値を持つクラスタペアを探索するクラスタリング手法である．タプル数 $|T|$ に対して最悪のケースで $O(|T|^3)$ の計算量を要する [5]．そのため，スケーラビリティに課題があり，大量のタプルを持つテーブル T を対象とした際に大きな計算時間を要する．また，各クラス間の併合においては RNR を極小化できるが，必ずしも最終的に生成されたクラスの RNR の小ささを保証するものではない．



(a) 初期状態　(b) 枝刈り　(c) 頂点の評価

図 1　類似度グラフ
Fig. 1　Similarity graph.

5. 効率的ノイズレスクラス生成手法

本章では，前章で述べた凝集型クラスタリングによる (ℓ_1, ℓ_2)-関係多様化の課題である効率性と，関係多様化の精度を，関係多様化の際に利用できるいくつかの性質を用いて改善する手法を提案する．

ここで，S_1 を前提部，S_2 を結論部とし，$v_i \in S_1$，$y_i \in S_2$ とする．まず，前提部のセンシティブ属性値ごとに，同一のセンシティブ属性値を持つタプルの集合を前提部タプル集合 $T[v_i]$ とする．ノイズレスクラスを生成するためには，ℓ_1 種類の $T[v_i]$ から同一の ℓ_2 種類の結論部値の集合を抽出してクラスをなす必要がある．

提案手法では，ノイズレスクラスを多数生成することで，テーブル全体の関係ノイズ比を低減させることを考える．共通の結論部を多く持つ前提部タプル集合からは多くのノイズレスクラスを生成できるため，結論部の共通性を基に ℓ_1 個の前提部タプル集合を抽出して，それらに属するタプル集合からノイズレスクラスを生成する．

以降では，5.1 節でノイズレスクラスをなしうるタプル集合が持つ性質について議論し，5.2 節でノイズレスクラス生成のアルゴリズムを提案する．

5.1 ノイズレスクラス生成の準備

本節では，ノイズレスクラスを効率的に生成するために，ノイズレスクラスをなしうるタプル集合が持つ性質について考える．

5.1.1 関係ベクトルと類似度グラフ

まず，前提部タプル集合ごとに結論部のセンシティブ属性値の頻度をベクトル形式で表現する．

$$\boldsymbol{v_i} = (v_{i,1}, \ldots, v_{i,|S_2|})^T \tag{14}$$

ここで，$t.S_1 = v_i$，$t.S_2 = y_j$ であるタプルの頻度を $f_{i,j}$ とし，$v_{i,j} = f_{i,j} / \sum_{\ell=1}^{|S_2|} f_{i,\ell}$ とする．$v_{i,j}$ は $\sum_{\ell=1}^{|S_2|} f_{i,\ell}$ によって正規化された $t.S_1 = v_i$，$t.S_2 = y_j$ であるタプルの頻度を表す．以降，$\boldsymbol{v_i}$ を関係ベクトルと呼ぶ．

ベクトルの類似度は，ベクトル間の内積（コサイン類似度）等によって求めることができる．ベクトル間の内積は，共通要素の出現パターンの類似性を表す．一方，出現パターンが類似する関係ベクトルのタプル群からは，同一の結論部を持つタプルが多く存在する可能性があり，ノイズレスクラスを多数生成できる可能性がある．そこで，2 つの関係ベクトルのタプル群からノイズレスクラスを生成できる可能性の高さを，関係ベクトル間の内積によって求める．

さらに，ノイズレスクラスを 1 つ以上生成するためには，ℓ_1 個の関係ベクトル間で共通の結論部が ℓ_2 個以上存在する必要がある．よって，2 つの関係ベクトル間でも共通の結論部が ℓ_2 個以上存在する必要がある．

以上より，命題 1 が導かれる．

命題 1　(k, ℓ_2)-関係多様性を満たすノイズレスクラスを生成可能な前提部タプル集合 $T[u_1], \ldots, T[u_k]$ は，共通の結論部値を ℓ_2 以上持つ（$cmn(v_1, \ldots, v_k) \geq \ell_2$）．

ここで，2 つの関係ベクトルに共通の結論部値の種類数を cmn とする（式 (15)）．

$$cmn(\boldsymbol{v_p}, \boldsymbol{v_q}) = \sum_{y_j \in S_2} sgn(v_{p,j} v_{q,j}) \tag{15}$$

これまでの議論から，関係ベクトル間の類似性（ノイズレスクラスの生成可能性）を以下の式で定義する．

$$sim(\boldsymbol{v_p}, \boldsymbol{v_q}) = \begin{cases} \boldsymbol{v_p} \cdot \boldsymbol{v_q} & (cmn(\boldsymbol{v_p}, \boldsymbol{v_q}) \geq \ell_2) \\ 0 & (otherwise) \end{cases} \tag{16}$$

(ℓ_1, ℓ_2)-関係多様なノイズレスクラスを生成可能か否かを判定するためには，ℓ_1 個の関係ベクトル間の類似度を計算する必要がある．この類似度計算は，ℓ_1 の大きさに比例して対象とする関係ベクトルの組合せが大きくなり，計算のコストが大きくなることが予想される．全探索によって類似度計算した場合には $O(|S_1|^{\ell_1})$ の計算量を要し，ℓ_1 の大きさに比例して計算コストが増大する．

ここで，式 (16) の関係ベクトル間の類似度からは，類似度行列を生成できる．この類似度行列に基づいて，各関係ベクトルを頂点とし，類似度が 0 より大きい頂点にエッジを張ることで，関係ベクトル間の類似度を表す無向グラフ $G := (V, E)$ が生成できる（図 1 (a)）．ここで，頂点集合 V は関係ベクトル $\boldsymbol{v_i}$ の集合であり，エッジ集合 E は関係ベクトル間の類似関係を表す $\{\boldsymbol{v_i}, \boldsymbol{v_j}\}$ の集合である．エッ

ジで接続された関係ベクトルどうしは，$(2, \ell_2)$-関係多様なノイズレスクラス生成の候補である．

類似度グラフ上で，前述の関係ベクトル間の類似度計算について考える．ℓ_1 個の関係ベクトル間の類似度計算は，各頂点を始点とした深さ優先探索によって実現することが可能である．この類似度計算のための深さ優先探索には，深さ優先探索が $O(|V|+|E|)$ の計算量を有することから，$O(|V|(|V|+|E|))$ の計算量を要すると考えられる．前述の全探索と比較すると計算コストは小さいが，類似度グラフが更新されるたびに深さ優先探索を実施することを考えると，依然として計算コストが高い．

提案手法では，本項で導入した関係ベクトル，類似度グラフを前提とし，2 頂点間の類似度をベースとしたヒューリスティックな手法でノイズレスクラス生成を行う．そのために，5.1.2 項では，ヒューリスティックな手法を実現するためのいくつかの性質について考える．

5.1.2　類似度グラフの枝刈り

本項では，ノイズレスクラスをなす関係ベクトル集合が持つ性質について考える．これによって類似度グラフを枝刈りし，ノイズレスクラスをなしえない関係ベクトルをノイズレスクラス生成の候補から除外する．

ノイズレスクラスを成すためには結論部の共起が必要である．そのため，類似度グラフ上において，各頂点は $\ell_1 - 1$ 個の他の頂点と接続されている必要がある．

命題 2　(ℓ_1, ℓ_2)-関係多様性を満たすノイズレスクラスを生成可能な頂点 $\boldsymbol{v} \in V$ は，エッジ数が ℓ_1-1 以上の頂点だけである．

命題 2 より，グラフ G からノイズレスクラスになりえない頂点を簡単に枝刈りすることができる．

さらに，ノイズレスクラスをなすためには，ℓ_1 個の関係ベクトルが互いに共通の結論部を ℓ_2 個持つ必要がある．そのため，類似度グラフ上において，少なくとも互いにエッジが張られていることが必要であり，互いに接続された ℓ_1 個の頂点だけがノイズレスクラスをなしうる．言い換えると，頂点集合 $\forall \boldsymbol{v_1}, \ldots, \boldsymbol{v_k} \in V$ からノイズレスクラスを生成するには，$\forall \boldsymbol{v_i}, \boldsymbol{v_j} (i \neq j) \in \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}\}$ にエッジが張られたクリークをなすことで必要である．

命題 3　(ℓ_1, ℓ_2)-関係多様性を満たすノイズレスクラスが生成可能な頂点の集合 $V' \subset V$ は，$|V'| \geq \ell_1$ であり，$\forall \boldsymbol{v} \in V'$ で ℓ_1 次のクリークをなす V' だけである．

命題 3 を用いると，多くのエッジを枝刈りできる可能性がある．しかしながら，特定のグラフから指定された大きさのクリークを探索する問題は NP 完全であることが知られている [6]．さらにノイズレスクラス生成によって関係ベクトル間の類似度が変化し，グラフが更新されるたびにクリークの探索を行うことはリーズナブルでない．

本稿では，クリークの探索によるノイズレスクラス生成は行わず，命題 2 よってノイズレスクラスをなしうる関係ベクトルを絞り込んだうえで，2 頂点間の類似性によっての高い頂点集合からノイズレスクラスを生成する．

5.2　ノイズレスクラス生成アルゴリズム

提案手法では，ノイズレスクラスを優先的に生成し，関係ノイズ比の低減を図る．

提案手法は以下の手順のアルゴリズムによってノイズレスクラスを生成する．

(1) 関係ベクトルの生成と類似度グラフの生成．
(2) ノイズレスクラス生成の対象とする関係ベクトル群を抽出（5.2.1 項）．
(3) 前提部を多様化するようにクラス併合し，$(\ell_1, 1)$-関係多様化（5.2.2 項）．
(4) 結論部を多様化するようにクラス併合し，(ℓ_1, ℓ_2)-関係多様化（5.2.3 項）．
(5) 類似度グラフの更新（5.2.4 項）．
(6) (2)〜(4) をノイズレスクラスが生成できなくなるまで繰り返す．
(7) ノイズレスクラスをなしていないタプル群を関係多様化クラスタリング（4 章）．

(1) では，5.1 節に議論した関係ベクトルと類似度グラフの生成および枝刈りを行う．(2) では，関係ベクトル間の類似性からノイズレスクラス生成の対象とする ℓ_1 個の関係ベクトルを生成する．次に (3) では，(2) で抽出した関係ベクトル群のタプルを用いて，同じ結論部を持ち，異なる前提部を持つ $(\ell_1, 1)$-関係多様化したクラスを生成する．(4) では，(3) で生成した $(\ell_1, 1)$-関係多様化したクラスを併合して (ℓ_1, ℓ_2)-関係多様性を満たすクラスを生成する．(5) では，ノイズレスクラスを形成したタプルに関する情報を関係ベクトル，類似度グラフに反映し，これらの更新を行う．最後に，ノイズレスクラスを形成できなかったタプルについては，4 章に示した関係多様化クラスタリングによって関係多様化を行う（類似度には $DGRL$ を用いる）．

5.2.1　ノイズレスクラス生成の対象データ選択

本項では，ノイズレスクラスを生成しうるタプルの集合を，5.1 節で導いた性質等から選択する．

まず，命題 1 と命題 2 によってグラフ G を枝刈りし，ノイズレスクラスを生成しうる関係ベクトルを絞り込む．

次に，頂点 $\boldsymbol{v}$ にとって最良な併合対象である隣接頂点の選択を行う．ここで，$\boldsymbol{v}$ の隣接頂点の集合を $n_G(\boldsymbol{v})$ とする．このとき，$\boldsymbol{v}$ と類似度の高い隣接頂点 $\boldsymbol{u_1} \in n_G(\boldsymbol{v})$ はノイズレスクラスを生成できる可能性も高い．そこで，$\boldsymbol{v}$ との類似度が高い上位 ℓ_1-1 個の頂点の集合を $\boldsymbol{v}$ のノイズレスクラス生成候補とする．

続いて，どの頂点を中心としたノイズレスクラス生成を行うべきかを判断するために，各頂点がどれだけノイズレスクラス生成に適しているかを評価する．この評価では，

$\boldsymbol{v}$ を隣接頂点の類似度の積によって評価する．ここで，頂点 $\boldsymbol{v}$ の隣接頂点集合 $n_G(\boldsymbol{v})$ のうち，$\boldsymbol{v}$ との類似度が高い上位 ℓ_1-1 頂点の集合を $n^*_G(\boldsymbol{v})$ とする．頂点 $\boldsymbol{v}$ の評価値 $\pi(\boldsymbol{v})$ を式 (17) のように定義する．

$$\pi(\boldsymbol{v}) = \prod_{\boldsymbol{u} \in n^*_G(\boldsymbol{v})} sim(\boldsymbol{v}, \boldsymbol{u}) \tag{17}$$

$\pi(\boldsymbol{v})$ が最大の頂点 $\boldsymbol{v_{max}}$ とノイズレスクラス生成の基準とし，$\boldsymbol{v_{max}}$ と $\boldsymbol{u} \in n^*_G(\boldsymbol{v_{max}})$ の頂点集合 V' からノイズレスクラスを生成する．

5.2.2　前提部の多様化

ノイズレス生成を行うための候補となる関係ベクトル群を抽出したら，それらを用いてノイズレスクラスを生成する．まず，結論部に基づいて，ノイズのない (ℓ_1, 1)-関係多様性を満たすクラスを生成する．

関係ベクトル $\boldsymbol{v'_1}, \ldots, \boldsymbol{v'_{\ell_1}} \in V'$ に関するタプル集合を用いて，ノイズレスクラスの生成を行う．このとき，同一の結論部値と，異なる前提部値を持つタプル群によって (ℓ_1, 1)-関係多様性を満たすクラスの生成を行う．

まず，$T[\boldsymbol{v'_1}], \ldots, T[\boldsymbol{v'_{\ell_1}}]$ から任意の結論部値 y_j を持つタプルを 1 つずつ抽出し，クラスを生成する．同一の結論部値 y_j を持つクラスの集合を $C(y_j)$ とする．この操作をすべての $T[\boldsymbol{v'_1}], \ldots, T[\boldsymbol{v'_{\ell_1}}]$ から y を持つタプルを 1 つずつ抽出できなくなるまで繰り返す．

5.2.3　結論部の多様化

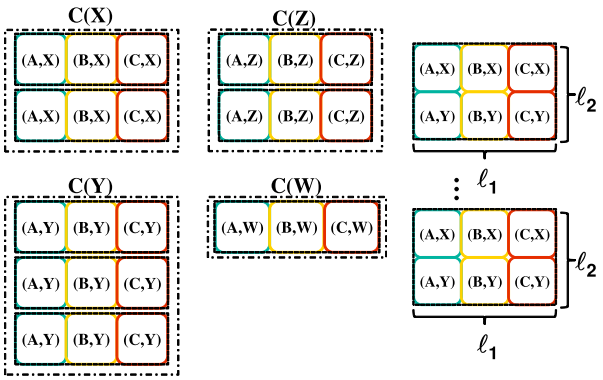
次に，結論部に基づくクラス併合で (ℓ_1, 1)-関係多様化したクラス群を，前提部の同一性に基づいて併合することで，(ℓ_1, ℓ_2)-関係多様化する．

ここで，結論部値 y ごとに生成できたクラス数を $f(y)$ とする．$f(y)$ が大きい順に上位 ℓ_2 個の結論部値を選択する．ここで選択された結論部値の集合を Y とする．$y \in Y$ の $C(y)$ からクラスを 1 つずつ選択し，1 つのクラスへ併合する．これによって (ℓ_1, ℓ_2)-関係多様性を充足するノイズレスクラスが生成される．

例 2　**図 2** (a) と図 2 (b) は，(3, 2)-関係多様化の例を示している．まず結論部値 X，Y，Z，W ごとにクラス集合 $C(X)$, $C(Y)$, $C(Z)$, $C(W)$ を生成する（図 2 (a)）．次に，頻度の高い上位 ℓ_2=2 件のクラス集合 $C(Y)$, $C(X)$ から 1 つずつクラスを抽出して，抽出したクラス群を 1 つのクラスに併合する（図 2 (b)）．この操作を，上位 ℓ_2=2 件のクラス集合から 1 つずつクラスが抽出できなくなるまで繰り返す．

5.2.4　類似度グラフの更新

ノイズレスクラスの生成に関与した結論部値の $f(y)$ を更新し，$f(y)$ が 1 以上の結論部値が ℓ_2 個以上存在する間，ノイズレスクラス生成を繰り返す．$f(y)$ が 1 以上の結論部値が ℓ_2 未満になったら，V' によるノイズレスクラス生成を終了し，$\forall \boldsymbol{v'} \in V'$ の関係ベクトルを更新し，$\boldsymbol{v'}$ と $\forall \boldsymbol{v} \in V$ の類似度を再計算する．再計算した類似度に基づいて類似度グラフ G の頂点とエッジを枝刈りする．G から頂点がなくなるまでノイズレスクラスの生成を繰り返す．



(a) 結論部に基づく (ℓ_1, 1)-関係多様化　(b) 前提部に基づく (ℓ_1, ℓ_2)-関係多様化

図 2　ノイズレスクラス生成
Fig. 2　Noiseless class grouping.

6.　評価

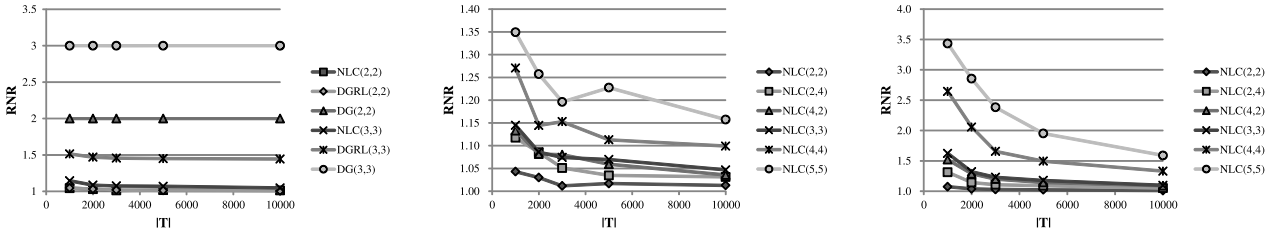
本稿の提案手法の有効性を評価するために，評価実験を行った．

効率的ノイズレスクラス生成手法を NLC（5.2 節），関係ノイズ比を考慮した関係多様化クラスタリングによる手法（4.2 節）を DGRL，関係ノイズ比を考慮しないナイーブな関係多様化クラスタリングによる手法（4.1 節）を DG として，これら 3 手法の関係ノイズ比，ノイズレスクラスの割合，関係多様化の計算時間を比較する．NLC，DGRL，DG は，本稿に記載した各手法を実装し，評価に利用した．

6.1　評価環境

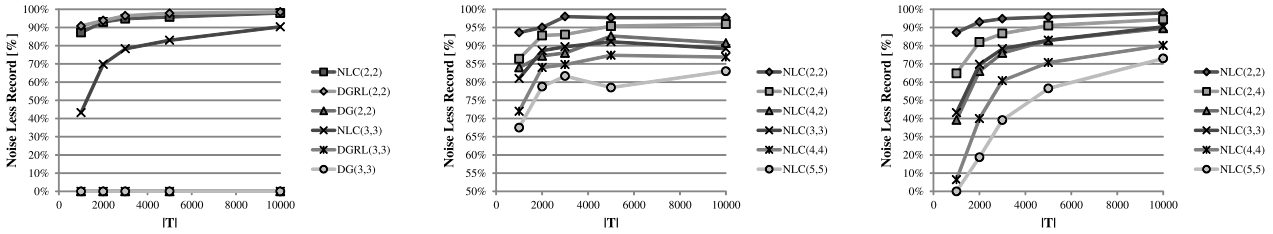
評価用のデータセットとして，2 種類の人工データを生成した．生成した人工データは，2 つのセンシティブ属性値にそれぞれ 10 種類の値を持つデータセット SA10 と，50 種類の値を持つデータセット SA50 であり，各タプルにはランダムに値を割り当てることで生成した．本評価では 1,000，2,000，3,000，5,000，10,000 件のタプルを持つ SA10，SA50 を生成した．なお，本評価では，関係多様化によるノイズの発生をどれだけ抑制できるかを純粋に評価するために，準識別子を含まない人工データを対象とした．

本評価では，仮想マシン上で評価を行った．評価で用いた仮想マシンは，4 コア CPU と，32 GB のメモリ，120 GB のディスクを持つ．仮想マシンのホストは，12 コア（24 スレッド）の 2.4 GHz の CPU と，192 GB のメモリ，6 TB のディスクを持つ．仮想マシン上で Java 言語（Java 1.6.0_32）によって実装した．匿名化対象のデータセットは PostgreSQL（PostgreSQL 8.4.13）に格納し，匿名化結果も PostgreSQL に格納する．



(a) 手法の比較　(b) SA10　(c) SA50

図 3　関係ノイズ比の平均値
Fig. 3　Average RNR.



(a) 手法の比較　(b) SA10　(c) SA50

図 4　ノイズレスクラスの割合
Fig. 4　Proportion of noiseless classes.

6.2　評価項目と結果

6.2.1　関係ノイズ比

3.1 節で導入した関係ノイズ比 RNR を用いて，関係多様化したテーブルのクオリティを評価する．ここでは，クラスの RNR の平均値を評価した．**図 3** (a) は，(2, 2)，(3, 3)-関係多様性を充足させた 3 手法の RNR をデータセットサイズごとにプロットして示している．関係多様化の対象としたデータセットは SA10 である．

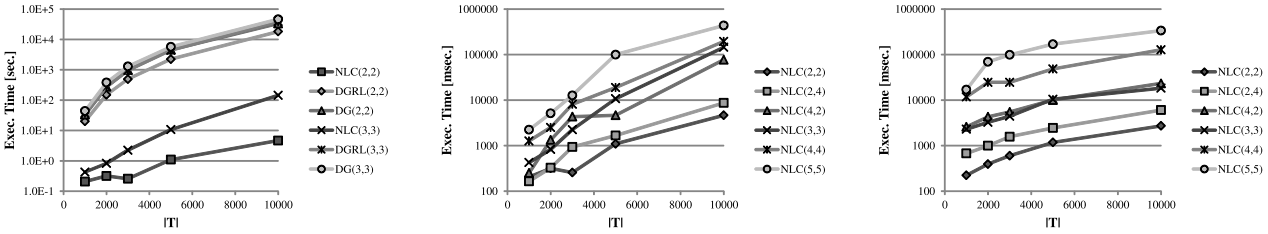
提案手法の NLC は，(2, 2)-関係多様化，(3, 3)-関係多様化において，RNR が 1 に近い値であり，関係多様化によるノイズの混入が非常に少ない．DGRL は，(2, 2)-関係多様化においてはノイズの混入率は非常に少ないが，(3, 3)-関係多様化の際にはノイズの混入によって本来の関係数の 1.5 倍程度の関係数に見えてしまう．RNR を考慮しないナイーブな手法である DG はノイズの割合が非常に多く，本来観測されうる関係を発見することが困難になってしまっている．以上より，クラス単位での局所的な関係ノイズ比においては，提案手法を低く抑えることができたといえる．

次に提案手法 NLC について，さらに詳細な評価を行った結果について示す．図 3 (b) は SA10 に対して，(2, 2)，(2, 4)，(4, 2)，(3, 3)，(4, 4)，(5, 5)-関係多様性をそれぞれ充足させた場合の RNR を示している．充足すべき多様性が小さいほど，RNR は小さい．(2, 4)，(4, 2)，(3, 3)-関係多様化結果を比較すると，(3, 3)-関係多様化した場合が最も RNR が高い．これは，ノイズレスクラス生成に必要となる最小レコード数が，(2, 4)，(4, 2)-関係多様化の場合は 8 であるのに対して，(3, 3)-関係多様化の場合は 9 であり，ノイズレスクラスを生成するためにより多くの種類の関係を持つレコードを必要とし，ノイズレスクラスの実現の困難さが高いためと考えられる．図 3 (c) は SA50 に対して同様に関係多様化をした際の RNR を示している．図 3 (b) と図 3 (c) を比較すると，属性値の種類数が多い SA50 を対象とした図 3 (c) の結果が明らかに RNR が高い．同一のレコード数である際には，属性値の種類数が多いほど同一の属性値を持つレコードが少なくなり，ノイズレスクラスを生成することが困難になるためと考えられる．

6.2.2　ノイズレスレコードの割合

次に関係多様化テーブル中のノイズレスクラスに属するレコード（ノイズレスレコード）の割合を評価した．**図 4** (a)，図 4 (b)，図 4 (c) には，関係多様化テーブルのレコードのうち，ノイズレスレコードの割合を示している．

図 4 (a) は各手法のノイズレスレコードの割合を比較した結果である．提案手法 NLC は (2, 2)-関係多様化の際には，90%以上のレコードがノイズレスレコードである．また，提案手法のヒューリスティクスを用いない DGRL も (2, 2)-関係多様化において，85%以上のレコードがノイズレスレコードである．しかしながら，RNR を考慮しない DG はノイズレスレコードが存在しない．よって，RNR を考慮して関係多様化することで，多くのレコードのセンシティブ属性の二項関係を過度に曖昧化しないことが分かる．(3, 3)-関係多様化の際には，他の手法がほぼノイズレスレコードを生成できていないことに対して，提案手法 NLC は多くのノイズレスレコードを生成している．これは，より多くのレコードがノイズレスクラスをなすようなレコー



(a) 手法の比較　(b) SA10　(c) SA50

図 5　計算時間
Fig. 5　Execution time.

ドのグループ化処理を導入しているためと考えられる．

さらに，提案手法 NLC について，いくつかの関係多様性を充足させた場合のノイズレスレコードの割合を図 4 (b) と図 4 (c) に示す．充足すべき関係多様性が小さいほど，多くのノイズレスレコードが生成できていることが分かる．しかしながら，ノイズレスクラスを優先的に生成する NLC であっても，充足すべき関係多様性が高い場合には，ノイズレスクラスがほとんど生成できないことが分かる．特に，図 4 (c) に示した SA50 を対象とした場合には，(5, 5)-関係多様化においてノイズレスレコードが 1 つもない．よって，データの性質を棄損せずに関係多様化をしたい場合には，データセットの性質や規模によって ℓ_1 や ℓ_2 を調整する必要がある．

6.2.3　計算時間

関係多様化手法のスケーラビリティを評価するために，各手法の計算時間を計測した．**図 5** (a) は各手法の計算時間を示している．提案手法 NLC は他の手法と比較して 10 倍以上高速であることが分かる．また，データサイズの増加に対して計算量が比較的増大しやすい傾向にあることが分かる．DGRL と DG はほぼ同程度の計算時間である．これは，DGRL と DG は距離計算の方法以外は同一の凝集型クラスタリングによって実現しているためである．NLC は凝集型クラスタリングに先立って，ヒューリスティクスを用いてノイズレスクラスを生成する．このノイズレスクラスが非常に効果的であると考えられ，かつ，関係多様化のために必要となる類似度計算の対象数を大幅に減らすことによって，DGRL と DG と比較して，大幅に高速化が実現できていると推察される．

図 5 (b) と，図 5 (c) は SA10，SA50 それぞれに対して NLC で関係多様化した際の計算時間を詳細に示している．データ数が少ない場合，センシティブ属性の値の種類が少ない方が計算時間が短い．これは，ノイズレスクラスを多く生成し，凝集型クラスタリングのコストを低減させることで計算時間が短縮されると考えられるため，センシティブ属性値の種類数が多い SA50 において，時間を要したのではないかと推察される．同様に，充足すべき関係多様性が高いほどノイズレスクラス生成が困難であるため計算時間が長いのではないかと考えられる．

以上より，提案手法 NLC は高い効率性を有しながら，関係多様化による関係の曖昧化を抑止できることが示された．

7.　関連研究

ℓ-多様性 [2]，t-近接性 [7]，m-不変性 [8] 等の k-匿名性を発展させ，センシティブ属性の多様性等に基づく匿名性指標が提案されている．特に ℓ-多様化 [2] では，準識別子の組合せからユーザのセンシティブ属性値を ℓ 種類未満に絞り込めないように準識別子を加工する．また，k-匿名化を確率的手法へ拡張した Pk-匿名化 [9] も提案されている．

同一のセンシティブ属性ではあるが，同一データ主体の複数のセンシティブ属性値を扱うデータに時系列データがある．時系列データに対する匿名化方式は，主に移動軌跡に対する匿名化方式が研究されている [10], [11], [12], [13]．しかしながら，いずれも k-匿名性を拡張させた手法であり，センシティブ属性間の多様性を保証するものではない．

また，集合（トランザクション）形式の属性値に対する匿名化手法もいくつか提案されている [14], [15], [16], [17], [18]．いずれの手法もトランザクション形式の 1 つの属性に関して，アイテム集合の重複性に基づく匿名化手法である．さらに，あるアイテムを知識とした他のアイテムの推定確率を一定以下に抑えることによるプライバシ保護手法についても提案されている [15], [18]．

テーブルを分割することによって，関係を曖昧化し，属性値の推定を困難にする技術が提案されている [19], [20]．Aggarwal ら [19] は，センシティブな関係が分割されるようにテーブル分割を行い，論理的に異なる 2 つ以上のサーバに配置して，関係の再構築を抑止する手法を提案している．Jiang ら [20] は，関係従属性のある属性間の関係をテーブル分割によって曖昧化し，ℓ-多様性を充足させる手法を提案している．両手法ともに，属性間の関係をテーブル分割によって曖昧化することについては本稿と同じモチベーションであるが，本稿のように，関係の曖昧化を抑止する仕組みが導入されていない．Jiang ら [20] の手法では，2 つ以上の属性間の関係の l-多様化を提案している．本稿の手法も二項関係をつなぎ合わせていくことで 2 つ以上の

属性間に対応することが可能であるが，ノイズの発生を抑止できるのは二項関係のみであるという制限がある．

8. おわりに

本稿では，2つのセンシティブ属性を持つレコード群に対するプライバシ保護の問題を扱った．特定のデータ主体に関するあるセンシティブ属性に関する知識から他のセンシティブ属性値が特定されるプライバシ侵害に対して，センシティブ属性間の関係多様化を提案した．さらに，関係多様化によって生じる関係の曖昧化を抑止しつつかつ，ナイーブなクラスタリング手法よりも効率性の高い関係多様化手法を提案した．評価実験では，提案手法がセンシティブ属性間の関係多様性を小さな曖昧化で実現でき，高い効率性を有することを示した．今後の課題として，従来のデータ匿名化で扱うような準識別子を属性として持つ場合の関係多様化方式の提案があげられる．また，本稿の関係ノイズ比は，クラス単位の局所的な視点に基づくものであり，テーブル全体の関係ノイズ比を全域的な視点での評価したものではない．全域的な視点での関係ノイズ比の定義も今後の課題の1つである．

参考文献

[1] Sweeney, L.: *k*-anonymity: A Model for Protecting Privacy, *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, Vol.10, No.5, pp.555–570 (2002).

[2] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: ℓ-Diversity: Privacy Beyond k-Anonymity, *ACM Trans. Knowledge Discovery from Data* (*TKDD*) (2007).

[3] Meyerson, A. and Williams, R.: On the Complexity of Optimal K-Anonymity, *Proc. PODS 2004*, pp.223–228 (2004).

[4] Day, W. and Edelsbrunner, H.: Efficient algorithms for agglomerative hierarchical clustering methods, *Journal of Classification*, Vol.1, No.1, pp.7–24 (1984).

[5] Olson, C.F.: Parallel Algorithms for Hierarchical Clustering, *Parallel Computing*, Vol.21, No.8, pp.1313–1325 (1995).

[6] Karp, R.M.: *Reducibility among Combinatorial Problems*, pp.85–103, Springer US (1972).

[7] Li, N., Li, T. and Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity, *Proc. ICDE 2007*, pp.106–115 (2007).

[8] Xiao, X. and Tao, Y.: m-invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets, *Proc. SIGMOD 2007*, pp.689–700 (2007).

[9] 五十嵐大，千田浩司，高橋克巳：k-匿名性の確率的指標への拡張とその応用例，*CSS 2009* (2009).

[10] Abul, O., Bonchi, F. and Nanni, M.: Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases, *Proc. ICDE 2008*, pp.376–385 (2008).

[11] Terrovitis, M. and Mamoulis, N.: Privacy Preservation in the Publication of Trajectories, *Proc. MDM 2008*, pp.65–72 (2008).

[12] Nergiz, M.E., Atozori, M., Saygin, Y. and Güç, B.: Towards Trajectory Anonymization: A Generalization-Based Approach, *Trans. Data Privacy*, Vol.2, pp.47–75 (2009).

[13] Takahashi, T. and Miyakawa, S.: CMOA: Continuous Moving Object Anonymization, *Proc. IDEAS 2012*, pp.81–90 (2012).

[14] Terrovits, M., Mamoulis, N. and Kalnis, P.: Privacy Preserving Anonymization of Set-valued Data, *Proc. VLDB 2008* (2008).

[15] Xu, Y., Wang, K., Fu, A. and Yu, P.S.: Anonymizing Transaction Databases for Publication, *Proc. SIGKDD 2008* (2008).

[16] He, Y. and Naughton, F.: Anonymization of set-valued data via top-down, local generalization, *Proc. VLDB 2009* (2009).

[17] Liu, J. and Wang, K.: Anonymizing Transaction Data by Integrating Suppression and Generalization, *Proc. PAKDD 2010* (2010).

[18] Cao, J., Karras, P., Raissi, C. and Tan, K.L.: ρ-uncertainty: Inference-proof Transaction Anonymization, *Proc. VLDB 2010*, pp.1033–1044 (2010).

[19] Aggarwal, G., Bawa, M., Ganesan, P., Garcia-Molina, H., Kenthapadi, K., Motwani, R., Srivastava, U., Thomas, D. and Xu, Y.: Two Can Keep a Secret: A Distributed Architecture for Secure Database Services, *Proc. CIDR 2005* (2005).

[20] Jiang, X., Gao, J., Wang, T. and Yang, D.: Multiple Sensitive Association Protection in the Outsourced Database, *Proc. DASFAA 2010* (2010).

髙橋 翼　（正会員）

2008年筑波大学第三学群情報学類卒業．2010年同大学大学院システム情報工学研究科コンピュータサイエンス専攻博士前期課程修了．同年日本電気株式会社入社．2013年筑波大学大学院システム情報工学研究科博士後期課程入学．主としてパーソナル情報の利活用におけるプライバシ保護の研究に従事．日本データベース学会会員．

側高 幸治

1997年中央大学理工学部電気・電子工学科卒業．1999年同大学大学院理工学研究科電気電子工学専攻博士前期課程修了．同年日本電気株式会社入社．現在，クラウドシステム研究所主任．暗号，PKI，匿名化技術等，セキュリティ基盤技術の研究開発に従事．

竹之内 隆夫　（正会員）

2003年電気通信大学電気通信学部情報工学科卒業．2005年同大学大学院情報システム学研究科博士前期課程修了．同年日本電気株式会社入社．現在，クラウドシステム研究所主任．2013年電気通信大学大学院情報システム学研究科博士後期課程修了．博士（工学）．主としてパーソナル情報の利活用におけるプライバシ保護の研究に従事．電子情報通信学会会員．

森 拓也　（正会員）

1995年東京大学工学部船舶海洋工学科卒業．1997年同大学大学院工学系研究科情報工学専攻修士課程修了．同年日本電気株式会社入社．現在，クラウドシステム研究所主任研究員．主としてパーソナル情報の利活用におけるプライバシ保護の研究に従事．

（担当編集委員　乾 孝司）