Computer Security Symposium 2019
21 - 24 October 2019

次元削減を利用した匿名化データに対する有用性と安全性の評価

山添 貴哉[1,a)]　面 和成[1,2,b)]

概要：近年、機械学習や深層学習の発達により、データの利活用が盛んになった。それに伴い、企業・医療機関・金融機関などでは大規模なデータの収集がされており、ビックデータと呼ばれている。各機関で収集されたデータは、大規模である程有用な知見が得られると考えられるが、データには多くの個人情報が含まれているため、外部に公開し共有することは出来ない。さらに、データから個人情報を保護する手法として、k-匿名化などのデータ加工手法が提案されているが、高次元データの有用性を保ったまま匿名化することは困難であるとされている。そこで、高次元なデータの有用性を高く保ったまま匿名化するため、行列分解と匿名化を組み合わせた手法などが提案されている。本研究では、次元削減アルゴリズムを用いた匿名化手法を提案し、既存手法に対して安全性・有用性の比較評価を行い、提案手法の優位性を明らかにした。

キーワード：匿名化, プライバシー, 次元削減, 有用性, 安全性, 機械学習

Evaluation of Utility and Security for Anonimized Data Using Dimentional Reduction

Takaya Yamazoe[1,a)]　Kazumasa Omote[1,2,b)]

Abstract: In recent years, the use of data has become popular due to the development of machine learning and deep learning. Along with that, companies, medical institutions, financial institutions are collecting large-scale data, which is called big data. The data collected by each organization is considered to be useful knowledge as the scale is large, but since the data contains a lot of personal information, it cannot be disclosed to the outside and shared. Anonymization methods such as k-anonymization have been proposed as a method for protecting personal information from data, but it is difficult to anonymize while maintaining the usefulness of high-dimensional data. In order to anonymize while keeping the usefulness of data high, a technique combining matrix decomposition and anonymization has been proposed. In this study, we propose an anonymization method using dimensional reduction algorithm and compare existing methods with safety and usability.

Keywords: anonymization, privacy, reduction, utility, safety, machine learning

1. はじめに

ネットワーク上のデータ量の増加や機械学習・深層学習分野の発達によりデータの分析・利活用が盛んに行われるようになった。顧客の購買データは企業のマーケティングや推薦アルゴリズムに活用できる。また、医療機関が収集した患者の治療歴、薬品の投与量などのデータは新たな

[1] 筑波大学システム情報工学研究科リスク工学専攻
Systems and Information Engineering, University of Tsukuba Graduate School
[2] 情報通信研究機構
National Institute of Information and Communications Technology
[a)] s1920600@s.tsukuba.ac.jp
[b)] omote@risk.tsukuba.ac.jp

© 2019 Information Processing Society of Japan

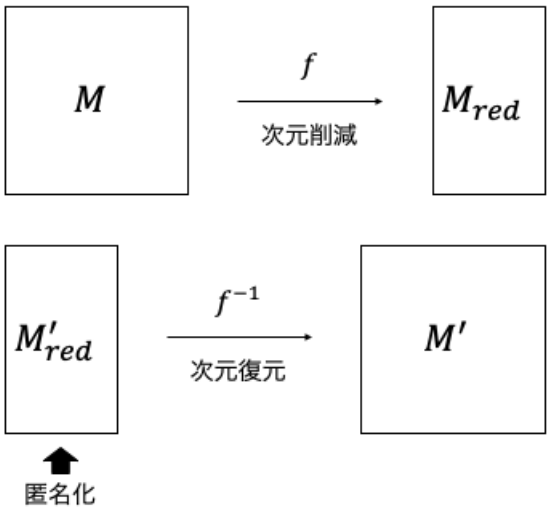

図 1 次元削減アルゴリズムと匿名化を組み合わせた手法

患者の診断補助に活用する事ができる。その他にも、位置情報やアクセスログなど、データを利活用することで、効果的な企業戦略を立てることや単純作業の効率化などが期待されている。この様に、データを利活用するためには、より多くのデータを収集する事が必要である。

しかし、データ利活用の有用性の反面、データ利活用にはプライバシーの問題が潜んでいる。購買履歴や治療歴、その他多くのデータには個人情報が含まれており、データ提供者のプライバシーを保護するために、データの取り扱いには慎重にならなければいけない。データ提供者のプライバシーが侵害された事例として、Netflix の事例がある。Netflix は、推薦アルゴリズムのコンペティションを開催するため、一部のユーザのデータを公開した。公開されたデータからはユーザを一意に特定する識別子は取り除かれていた。しかし、Narayanan ら [1] は特定の条件のもとで、攻撃者が公開データからユーザを一意に特定できることを統計的に示した。

データを利活用するだけの多量で多様なデータを 1 つの機関で収集することは難しく、複数の機関で収集したデータを 1 つに集約して分析する事が求められる。しかし上述の様に、データ公開にはプライバシーの問題が潜むため、データを無加工の状態で外部に公開することは避けるべきである。データを集約する際に、データを加工することで、データ提供者のプライバシーを保護する手法をデータの匿名化と呼び、代表的な手法に k-匿名化やノイズ付与などがある。しかし、代表的な手法の 1 つである k-匿名化は高次元データの有用性を高く保ったまま匿名化が困難であるとされており [3]、このことは次元の呪いと呼ばれている。そこで、高次元データの有用性を保ちつつ、データを匿名化する手法として、行列分解と k-匿名化やノイズ付与を組み合わせる手法が提案されている [6][7]。これらの手法は、行列分解で高次元データを低次元に変換することで、データの次元の呪いに対処している。

我々は、次元の呪いを克服するため、行列分解ではなく、次元削減アルゴリズムを用いる匿名化手法を提案する。具体的には、主成分分析や局所性保存射影、オートエンコーダーなどの次元削減アルゴリズムと k-匿名化やノイズ付与を組み合わせた手法を提案し、既存の匿名化手法や行列分解を用いた匿名化手法と安全性・有用性の観点で比較する。提案手法のデータ匿名化の流れを**図 1** に示す。

本研究の貢献は以下の通りである。

- 主成分分析 (PCA)、局所性保存射影 (LPP)、Autoencoder の 3 つの次元削減アルゴリズムを用いた新たな匿名化手法を提案する。

- 提案した匿名化手法と、既存の匿名化手法を安全性・有用性の 2 つの観点で定量的に評価し、次元削減アルゴリズムと匿名化アルゴリズムを組み合わせることで、匿名化データの安全性が向上すること、機械学習におけるデータの有用性が向上することを明らかにした。

2. 関連研究

2.1 k-匿名化

k-匿名化は、データベース中の少なくとも k 件以上のデータが、ある準識別子の中で同じデータを持つ様にデータを変換する匿名化手法である [2]。よって、そのデータの持ち主が k 人未満に絞り込めないという直感的な指標として扱われてきた。k-匿名性を論じる場合、一般的には各属性や組み合わせを識別子、準識別子とその他の属性に分類する。識別子はデータの持ち主を個人に特定できる属性、準識別子とは複数の組み合わせによってデータの持ち主を個人に特定できる属性のことを指す。識別子の例としては個人 ID など、準識別子の例としては名前や年齢などがあげられる。k-匿名化法では k-匿名性に基づき、データの準識別子からそのデータの持ち主を k 人以下に絞り込めないように、準識別子を削除、加工することで個人のプライバシーを保護する。しかし、扱うデータが高次元になると、データベース中のレコード間の距離が急激に大きくなってしまう。そのため k-匿名化は高次元のデータに対して有用性を高く保ったまま匿名化することが困難であることが示されており [3]、このことは次元の呪いと呼ばれている。本研究では、次元の呪いの問題に対処するため、次元削減アルゴリズムでデータを低次元に変換してから k-匿名化を行うことによって、有用性の高い匿名化データを得られることを示す。

2.2 行列分解匿名化

近年盛んになってきた機械学習やデータ分析には高次元データを用いることが多い。しかし、上述した様に、高次

元データの有用性を高く保ったまま匿名化データを生成することは困難であることが示されている [3]。そのため、本研究以外にも高次元データの有用性を保持しながら匿名化する手法が提案されている。その 1 つとして行列分解を用いた手法がある [6][7]。行列分解とは、$M \in \mathbb{R}^{n \times m}$ を 2 つの行列である $U \in \mathbb{R}^{n \times r}$ と $V \in \mathbb{R}^{m \times r}$ に分解する手法である。行列 $X = UV^{\mathrm{T}}$ は M の近似になっており、ランク r はその正確性を指定するパラメータである。Mimoto ら [7] は、行列分解と k-匿名化、行列分解とノイズ付与を組み合わせて高次元データの有用性を高く保ったまま匿名化できることを示した。しかし、Mimoto らの用いた非負値行列因子分解は負の値が含まれるデータには適用することが出来ない為、データの標準化などを行う事が想定されていないなどの問題点もある。本研究では、データの説明能力を高く保ったままデータを低次元に変換する事が出来る主成分分析などの次元削減アルゴリズムを利用することによって、行列分解と匿名化を組み合わせて得られた匿名化データよりも、有用性の高い匿名化データを生成できることを示す。

3. 準備

3.1 次元削減アルゴリズム

本節では、匿名化手法に用いた次元削減アルゴリズムについて述べる。

3.1.1 主成分分析 (PCA)

PCA は学習データの X の分散が最大になる方向への線形変換を求めることによって、データの次元を削減する手法である。次元削減後の主成分軸は、学習データの共分散行列の固有ベクトルを求めることによって得られる。こうして得られた主成分は学習データの説明能力を多く保持していると考えられる。学習データの次元が大きい場合、データの解釈が困難になるが、データの説明能力を多く保持した低次元の主成分軸にデータを射影することで、データの解釈性を高める事が出来る。このように、PCA はデータの解釈性を高めるために用いられる手法であるが、本研究では有用性を保ったまま次元圧縮が出来る PCA の性質に着目し、匿名化手法への応用を提案する。

3.1.2 局所性保存射影 (LPP)

LPP[4] は PCA と同じく、データを低次元空間に射影するための線形変換を生成し、高次元データの低次元空間での写像を得る手法である。PCA との違いは、射影の際に類似度行列を用いることによって、局所でのクラスター構造を保持したまま低次元空間写像を得られる点にある。PCA と同様、本来はデータの解釈性を高めるために用いられる手法である。

3.1.3 Autoencoder

Autoencoder はニューラルネットワークを利用した次元圧縮アルゴリズムである [5]。ニューラルネットワークとは、入力層、中間層、出力層で構成される。通常のニューラルネットワークでは、入力層に学習データを入力し、出力層から得られる出力と正解データを比較することによってネットワーク全体を学習させる。Autoencoder では、正解データを学習データとすることで、出力層の出力が入力層への入力と等しくなるようにネットワークを学習させる。この様にネットワークを学習させることで、入力層と出力層よりも次元の低い中間層から、学習データの抽象化データを得る事が出来る。本研究では、Autoencoder の抽象化データを得るための線形変換を利用することによって匿名化データを生成する。

3.2 評価方法

本節では、本研究で用いる匿名化データの評価方法を示す。本研究では、匿名化データを安全性と有用性の 2 つの指標を用いて評価する。以下に安全性と有用性の評価方法を詳細に記載する。

3.2.1 安全性

安全性の評価には Re-Identification Attack[7] を用いる。Re-Identification Attack とは、攻撃者が匿名化前の元データと匿名化データを持つとき、2 つのデータセットのレコードを対応づける攻撃である。レコードの対応付けは最適割当問題と考える事ができるので、攻撃者が元データと匿名化データのレコード間の最適割当問題を解くという想定で、最適割当問題で正確に対応づけができた割合 (Accuracy) で匿名化データの安全性を評価する。

3.2.2 有用性

匿名化データの有用性は、データの利用方法に適した手法を選択して評価すべきである。本研究では、近年の機械学習分野の躍進に合わせ、匿名化データが機械学習に用いられることを想定する。そのため、本研究では三本ら [7] らの有用性の評価手法を採用する。具体的には、元データを用いて学習させたモデルの F 値を F_{ori}、匿名化データを用いて学習させたモデルの F 値を F_{ano} と置き、匿名化データの有用性を以下の式で評価する。

$$Uti(A(M)) = \frac{F_{ano}}{F_{ori}} \tag{1}$$

4. 提案手法

本章では、次元削減アルゴリズムを利用したデータ匿名化手法を提案する。次元削減アルゴリズムは与えられた多次元のデータから得られる写像 f を用いて、低次元に圧縮する事が出来る。本研究では、低次元へと圧縮したデータに対して k-匿名化などの匿名化手法を適用する。次に、圧縮した写像の逆写像 f^{-1} を用いて次元を復元して匿名化データを得る。よって、得られる匿名化データと元データの次元数は等しい。本研究で利用する次元削減アル

ゴリズムは 3.1 節で述べた主成分分析、局所性保存射影と Autoencoder である。以下に提案手法のアルゴリズムを示す。元のデータセットを M、次元削減アルゴリズムの写像を f、その逆写像を f^{-1}、匿名化アルゴリズムを A とする。写像 f を用いて次元削減データ $M_{reduction}$ をし、匿名化アルゴリズム A を用いて $M_{reduction}$ を匿名化し、$M^{'}_{reduction}$ を得る。その後に次元削減アルゴリズムの逆写像 f^{-1} を用いて元データと同じ次元数の匿名化データ $M^{'}$ を得る。提案手法のアルゴリズムを Algorithm1 に示す。

Algorithm 1 Anonymization algorithm

Require: dataset M, reduction algorithm f
$M_{reduction} = f(M)$
$M^{'}_{reduction} = A(M_{reduction})$
$M^{'} = f^{-1}(M^{'}_{reduction})$
return $M^{'}$

5. 実験と考察

本章では、本研究で行なった実験の結果を示す。実験内容は、3.2 節でも述べたとおり、匿名化データの安全性と有用性を評価するための実験である。

5.1 データセット

本研究では Otto Group の商品データセットを使用した。同グループはこのデータセットを公開し、2015 年に商品のカテゴリーを分類するコンペティション[*1] を開催している。このデータセットには 61,678 個の商品の情報が記録されている。それぞれの商品データは、各商品の持つ 93 個の特徴量とその商品が分類されるべきカテゴリーの情報で構成されている。商品は 9 つのカテゴリーのいずれかに分類される。本研究では、商品の特徴量をセンシティブな情報に見立て、匿名化処理と安全性・有用性の評価を行なった。

表 1 k-匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

k	3	10	50	100
acc	0.331	0.094	0.018	0.009

表 2 $A_k, f_{pca}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ k	3	10	50	100
30	0.246	0.059	0.011	0.004
50	0.251	0.059	0.012	0.006
70	0.253	0.065	0.008	0.005

[*1] Otto Group Product Classification Challenge https://www.kaggle.com/c/otto-group-product-classification-challenge

5.2 匿名化

本研究で行なった実験では、複数の匿名化データを生成した。提案手法には、次元削減アルゴリズムと匿名化アルゴリズムの 2 つのアルゴリズムが必要となる。今回の実験では、次元削減アルゴリズムに、主成分分析 (PCA)、局所性保存射影 (LPP)、Autoencoder の 3 つのアルゴリズムを用いた。これら 3 つの次元削減アルゴリズムをそれぞれ、f_{pca}、f_{lpp}、f_{ae} と表記する。これら 3 つの次元削減アルゴリズムに加え、既存手法である行列分解を f_{nmf} で表記する。オリジナルのデータセットは 93 次元の特徴量で表現されているが、次元削減アルゴリズムを用いて、次元を $d = 30, 50, 70$ に削減し、匿名化アルゴリズムを適用する。匿名化アルゴリズムには k-匿名化とノイズ付与 [8] を用いた。ノイズ付与とは、データにノイズを加えることでランダム化する単純な手法であり、本実験ではラプラス分布に従ったノイズ $\epsilon \sim Lap(0, 2\phi^2)$ を付与することで匿名化する。k-匿名化とノイズ付与はそれぞれ A_k、A_ϕ で表記する。以下の実験では、これらのアルゴリズムを組み合わせ、さらに匿名化アルゴリズムのパラメータ k と ϕ を変化させ、評価を行なった。オリジナルのデータセットを M とし、提案手法の匿名化アルゴリズムを $A, f(M)$ と表記する。

5.3 安全性評価

安全性の評価では、上述した次元削減アルゴリズムと匿名化アルゴリズムを組み合わせて生成した匿名化データを用いる。匿名化データの生成に用いたパラメータは k-匿名化の場合、$k = 3, 5, 10, 100$、ノイズ付与に関しては、組み合わせる次元削減アルゴリズムによって加えるべきノイズのパラメータが異なるため、各アルゴリズムに対して調整を行なった。これらの生成された匿名化データに対して、3.2.1 節で述べた Re-Identification Attack で評価した。次元削減アルゴリズムを用いず、オリジナル

表 3 $A_k, f_{lpp}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ k	3	10	50	100
30	0.081	0.005	0.003	0.003
50	0.146	0.020	0.001	0.001
70	0.210	0.029	0.004	0.001

表 4 $A_k, f_{ae}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ k	3	10	50	100
30	0.255	0.061	0.008	0.004
50	0.257	0.054	0.005	0.003
70	0.258	0.053	0.007	0.003

表 5　$A_k, f_{nmf}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ k	3	10	50	100
30	0.261	0.061	0.012	0.007
50	0.253	0.062	0.008	0.006
70	0.255	0.055	0.007	0.007

表 6　$A_\phi(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

ϕ	2.5	3.5	4.5	5.5	6.5
acc	0.529	0.163	0.075	0.03	0.015

表 7　$A_\phi, f_{pca}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ ϕ	1.0	1.5	2.5	3.5
30	0.737	0.492	0.235	0.146
50	0.851	0.641	0.346	0.182
70	0.878	0.693	0.406	0.227

表 8　$A_\phi, f_{lpp}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ ϕ	0.001	0.003	0.005	0.01
30	0.336	0.087	0.032	0.006
50	0.699	0.303	0.096	0.019
70	0.933	0.632	0.257	0.054

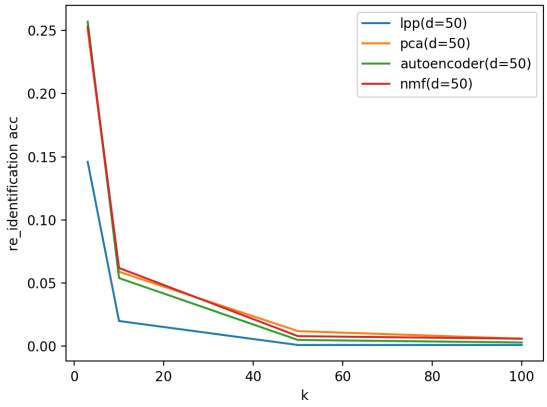

図 2　re-identification attack の成功率

のデータセットに匿名化アルゴリズムを適用した $A_k(M)$ に対する評価結果を**表 1** に示す。k-匿名化と次元削減アルゴリズム、行列分解を組み合わせた手法 $A_k, f_{pca}(M)$、$A_k, f_{lpp}(M)$、$A_k, f_{ae}(M)$、$A_k, f_{nmf}(M)$ に対する評価結果をそれぞれ**表 2**、**表 3**、**表 4**、**表 5** に示す。また、次元削減アルゴリズムと行列分解のパラメータを $d = 50$ として、匿名化アルゴリズム k を変化させた時の Re-Identification Attack の成功率の変化を**図 2** に示す。ノイズ付与と次元削減アルゴリズムを組み合わせて生成した匿名化データについても同じようにまとめた。ノイズ付与のみを用いて匿名化したデータの評価結果を**表 6** に、$A_\phi, f_{pca}(M)$、$A_\phi, f_{lpp}(M)$、$A_\phi, f_{ae}(M)$、$A_\phi, f_{nmf}(M)$ に対する評価結果を、**表 7**、**表 8**、**表 9**、**表 10** に示した。

k-匿名化と組み合わせたデータに関しては、図 2 より次元削減アルゴリズムを用いた場合も行列分解を用いた場合も、k-匿名化のみを適用した匿名化データより高い安全性を保証できていることが分かる。また、使用する次元削減アルゴリズムによって安全性に影響を及ぼすことが分かった。最も高い安全性を保証しているのは LPP を用いた匿名化データであるが、これは LPP の局所性のクラスター構造を保持できる性質と k-匿名化のアルゴリズムに親和性があったためと考えられる。ノイズ付与と組み合わせたデータに関しては、加えるノイズのパラメータを調整することによって、安全性が緩やかに変化する傾向にある。表 7 を見ると、パラメータの値が小さい時には k-匿名化よりも安全性が低いが、パラメータの値を徐々に大きくすることによって安全性を高くすることができる。

5.4　有用性評価

有用性の評価では、5.3 節で評価した安全性を考慮し、同水準の安全性を持つ匿名化データ同士で比較を行う。まずは、全ての匿名化データから安全性基準が、0.1 程度で同水準である匿名化データの有用性評価結果を**表 11** に示す。パラメータは、次元削減アルゴリズムに関しては、$f_{method=d}$ で表記しており、$method$ は次元削減に用いたアルゴリズムを、d は次元削減後のデータの次元を表している。匿名化アルゴリズムに関しては、$A_{method=p}$ で表記しており、$method$ は匿名化に用いたアルゴリズムを、p は匿名化に用いたパラメータを表している。表から Accuracy、F 値、Precision、Recall の指標で次元削減アルゴリズムと匿名化アルゴリズムを組み合わせた提案手法で生成した匿名化データの有用性が高いことが分かる。このことから、データの有用性を保ったまま次元圧縮ができる次元削減アルゴリズムは、匿名化の課題である次元の呪いの解決策として有効であると考えられる。

次に、次元削減アルゴリズムによる有用性の変化を検討するため、次元削減アルゴリズムと k-匿名化、次元削減アルゴリズムとノイズ付与を組み合わせた手法で、安全性が同水準の匿名化データを比較した結果を**表 12**、**表 13** に示す。この結果から、データの分散が大きくなるように次元圧縮をする PCA が最もデータの有用性を高く保ったままデータを匿名化することに適していると考えられる。

表 9　$A_{\phi}, f_{ae}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ ϕ	0.1	0.3	0.5	1.0
30	0.983	0.845	0.486	0.139
50	0.993	0.906	0.653	0.228
70	0.997	0.986	0.864	0.396

表 10　$A_{\phi}, f_{nmf}(M)$ で匿名化したデータへ Re-Identification Attack を実施した時の Accuracy

d \ ϕ	0.1	0.3	0.5	1.0
30	0.431	0.067	0.014	0.003
50	0.613	0.098	0.021	0.002
70	0.687	0.104	0.022	0.004

表 11　同水準の安全性を持つ匿名化データの有用性評価結果

Dataset	Accuracy	F measure	Precision	Recall
$A_{k=10}(M)$	0.487	0.353	0.438	0.469
$A_{k=10}, f_{pca=70}(M)$	0.739	0.577	0.655	**0.637**
$A_{k=3}, f_{lpp=30}(M)$	0.595	0.474	0.465	0.564
$A_{k=10}, f_{ae=30}(M)$	0.594	0.399	0.446	0.498
$A_{k=10}, f_{nmf=50}(M)$	0.583	0.417	0.511	0.514
$A_{\phi=4.5}(M)$	0.594	0.371	0.653	0.394
$A_{\phi=3.5}, f_{pca=30}(M)$	**0.747**	0.6	**0.702**	0.608
$A_{\phi=0.003}, f_{lpp=50}(M)$	0.571	0.455	0.447	0.516
$A_{\phi=1.0}, f_{ae=30}(M)$	0.702	**0.612**	0.592	**0.637**
$A_{\phi=0.3}, f_{nmf=50}(M)$	0.545	0.326	0.529	0.355

表 12　次元削減アルゴリズムによる有用性の変化 k-匿名化

Dataset	Accuracy	F measure	Precision	Recall
$A_{k=3}, f_{pca=30}(M)$	**0.733**	**0.581**	**0.6**	**0.637**
$A_{k=3}, f_{lpp=70}(M)$	0.593	0.389	0.42	0.518
$A_{k=3}, f_{ae=30}(M)$	0.653	0.545	0.522	0.596

表 13　次元削減アルゴリズムによる有用性の変化 ノイズ付与

Dataset	Accuracy	F measure	Precision	Recall
$A_{\phi=3.5}, f_{pca=70}(M)$	**0.733**	0.614	**0.764**	0.608
$A_{\phi=0.005}, f_{lpp=70}(M)$	0.521	0.424	0.407	0.508
$A_{\phi=1.0}, f_{ae=50}(M)$	0.72	**0.629**	0.604	**0.653**

6. おわりに

本研究では、データの匿名化の課題である次元の呪いに対処するため、次元削減アルゴリズムと匿名化アルゴリズムを組み合わせる手法を提案した。次元削減アルゴリズムでオリジナルのデータの有用性を保ったまま次元圧縮したデータを匿名化し、逆写像を用いて次元を復元することで、有用性の高い匿名化データを生成することができる。さらに、既存手法の行列分解と匿名化アルゴリズムを組み合わせた手法と比較実験を行い、提案手法が同水準の安全性で有用性が高いことを示した。

謝辞　この成果は、国立研究開発法人新エネルギー・産業技 術総合開発機構 (NEDO) の委託業務の結果得られたものです。

参考文献

[1] A.Narayanan, V.Shmatikov, Robust de-anonymization of large sparse datasets, In Proceedings of 2008 IEEE Symposium on Security and Privacy(SP), 2008

[2] S.Latanya, k-anonymity: a model for protecting privacy, International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, Volume 10, Issue 5, pp.557-570, 2002

[3] Charu C.Aggarwal, On k-Anonymity and the Curse of Dimensionality, Proceedings of the 31st international conference on Very large data bases, 2005

[4] X.He, P.Niyogi, Locally Preserving Projections, 16th International Conference on Neural Information Processing Systems, 2003

[5] G.E.Hinton, R.R.Salakhutodinov, Reducing the Dimensionality of Data with Neural Networks, Science, Vol313, 2006

[6] Satoshi.Hasegawa, Ryo.Kikuchi, Shogo.Masaki, Koki.Hamada, 行列分解を利用した確率的 k-匿名性を満たす高次元データ公開法,CSS,2016

[7] Tomoaki.Mimoto, Shinsaku.Kiyomoto, Seria.Hidano, Anirban.Basu, The Possibility of Matrix Decomposition as Anonymization and Evaluation for Time-sequence Data, Annual Conference on Privacy, Security and Trust, 2018

[8] J.J.Kim, A method for limiting disclosure in microdata based on random noise and transformation, in Proceedings of the section on survey research methods, American Statistical Association, 1986