SCIS 2022 2022 Symposium on Cryptography and Information Security
Osaka, Japan & Online, Jan. 18 – 21, 2022
The Institute of Electronics, Information and Communication Engineers

現実的な攻撃モデルを通した主成分分析のプライバシー的観点からの考察
A Study of the Privacy Perspective on Principal Componet Analysis via a Realistic Attack Model

山城 大海 * 面和成 *
Hiromi Yamashiro Kazumasa Omote

キーワード プライバシー, 主成分分析, 匿名化

あらまし

機械学習の隆盛に伴い，データセットに対するプライバシー保護が重要となっている．様々なプライバシーの定義やプライバシー保護手法が存在するが，基本的な方法としてデータに工夫されたノイズを付与することが挙げられる．しかし，ノイズはそのデータを用いた機械学習モデルの性能に影響する．プライバシー保護とデータの有用性の間にはトレードオフが存在する．

次元削減は，直接扱うことが困難もしくは非効率な高次元データを，その特徴を保ちつつ次元の小さなデータに変換する手法の総称である．主成分分析（PCA）はその 1 つであり，軸の分散を最大化しつつ次元削減を行う．PCA では，サンプルの分散共分散行列から，変換に用いる行列を計算する．

PCA をベースとしたプライバシー保護手法はよく研究されている．Chen らは次元削減を応用したプライバシー保護手法を提案し，その中で PCA を例として用いている [1]．Chen らの方法では，次元削減アルゴリズムで圧縮したデータに対しノイズを付与することでデータのプライバシーを向上させる．Chen らは PCA に対する攻撃モデルとして以下を用いている:

$$\mathbf{X} \Leftarrow \mathbf{W} \times \mathbf{T}^{\top} + mean. \tag{1}$$

$\mathbf{X}, \mathbf{W}, \mathbf{T}, mean$ はそれぞれ復元されたオリジナルデータ，PCA 後の圧縮されたデータ，変換行列，オリジナルデータの平均である．すなわち，Chen らのモデルは攻撃者が PCA に用いられた変換行列を入手できることを想定している．しかし，現実的には，攻撃者が変換行列を入手できるとは限らない．

本研究では，攻撃者の入手できる情報を制限した PCA に対する現実的な攻撃モデルを提案し，実験評価を行う．さらに，提案モデルの下で，従来なされてこなかった通常の PCA に対するプライバシー的観点からの考察を行う．

提案モデルにおいて，攻撃者は次の 2 段階を経る:

段階 1 補助情報から変換行列 $\mathbf{T}$ を推定する．

段階 2 推定された $\mathbf{T}$ を用いて式 (1) によりオリジナルデータを復元する．

補助情報とは，攻撃者の利用できる制限された情報である．実験では，補助情報としてオリジナルデータと同分布のサンプルを用いた．攻撃者は補助情報の分散共分散行列を計算し，それを用いて変換行列 $\mathbf{T}$ を推定する．実験の結果，$\mathbf{T}$ に補助情報から推定された行列を用いると，オリジナルの変換行列を用いたときに比べ，攻撃精度が著しく低下することが確認された．

これより，プライバシー保護手法を評価するにあたり，攻撃者に理想的な状況以外を想定することも必要であるといえる．また，現実的な仮定を置いた場合，ノイズ付与を行わない通常の PCA でも十分にプライバシーが保護される可能性が示唆される．

参考文献

[1] Chen, Omote, "A Privacy Preserving Scheme with Dimensionality Reduction for Distributed Machine Learning", AsiaJCIS 2021, pp.45-50, 2021.

* 筑波大学, 〒 305-8573 茨城県つくば市天王台 1-1-1, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 Japan.