社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

ランダムユニタリ変換により分散秘匿化された信号に対するLASSO解推定

坂東　幸浩†　仲地　孝之†　貴家　仁志††

† 日本電信電話株式会社
〒239-0847 神奈川県横須賀市光の丘 1-1
†† 首都大学東京　システムデザイン研究科
〒 191-0065 東京都日野市旭ヶ丘 6-6
E-mail: †{yukihiro.bandou.pe,takayuki.nakachi.pu}@hco.ntt.co.jp, ††kiya@tmu.ac.jp

あらまし　エッジ／クラウドコンピューティング上に集約されたビッグデータ解析が重要性を増す一方で、個人特定に繋がる可能性のあるデータの場合、プライバシー保護の観点から、データを取得した組織に閉じて利用される傾向にある。このため、十分なデータ量を確保できず、所望の分析精度が実現できない場合がある。そこで、本稿では、ランダムユニタリ変換により秘匿化されたデータに対して、Coordinate descent algorithm による LASSO 解の導出を通して分析モデルを構築する。さらに、秘匿化のスキームを分散した各拠点において個別に秘匿化する分散秘匿化に拡張し、分散秘匿化に対して、上記 LASSO 解の保全性に関する理論的保証を与える。本稿で提案する分散秘匿化により、プライバシー保護が必要なデータが分散取得された場合であっても、データの機密性は確保した上で、大規模なデータを利用した分析の高精度化が可能となる。
キーワード　LASSO、座標降下法、ランダムユニタリ変換

LASSO analysis for distributed encryption signal with random unitary transform

Yukihiro BANDOH†, Takayuki NAKACHI†, and HITOSHI KIYA††

† Nippon Telegraph and Telephone Corporation,
1-1 hikari-no-oka, Yokosuka, Kanagawa, Japan, 239-0847
†† Information and Communication Systems, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo, Japan, 191-0065
E-mail: †{yukihiro.bandou.pe,takayuki.nakachi.pu}@hco.ntt.co.jp, ††kiya@tmu.ac.jp

Abstract　Big-data analysis on edge/cloud becomes more important. However, when information may lead to personal identification, such information tend to be closed to its owners from the viewpoint of privacy protection. In such case, it is not possible to obtain enough data for the analysis. As a result, the desired analysis accuracy may not be achieved. In this paper, we construct an analysis model for data encrypted with the random unitary transform through deriving the LASSO solution of the encrypted data. The analytical model can derive the same LASSO solution as that of original data without encryption. The analytical model allows collaboration with a distributed encryption that encrypts data at each site independently. The collaboration enables us to improve the accuracy of analysis for distributed privacy-sensitive information.
Key words　LASSO, coordinate descent algorithm, random unitary transform

1. はじめに

近年、エッジ／クラウドコンピューティングはビッグデータ解析の計算リソースとして急速に普及している。その解析対象は、音声・映像等のメディア信号から商品取引情報等の経済データ、臨床結果等の医療データまで多岐に渡る。

しかし、取得データの個人特定に繋がる可能性のあるデータは、プライバシー保護の観点から、エッジ／クラウドコンピューティングの利用は制限される。例えば、臨床検査結果、購買履歴、移動経路などデータがあげられる。こうしたデータに対しては、データを取得した組織・機関に閉じて利用される。大量のユーザを抱え、所望の規模のデータを取得可能な場合は、問題ない。しかし、医療機関における臨床データのように、各機関で取得可能なデータ数が限られている場合、各機関に閉じた分析では、十分な分析精度を得られない場合がある。問題の原因は、取得されたデータが分散しており、集約できない点にある。

こうした問題を解決する方法の一つとして、データを暗号化した状態で計算可能な秘密計算が研究されている。秘密計算は、一般にマルチパーティプロトコルや準同型暗号に基づき実行される [1] [2] [3] [4] 。しかし、秘密計算は、除算の困難性、計算効率および計算精度に課題がある。このため、その適用は、ソーティング処理や幾つかの統計処理に限定されており、十分な普及にはいたっていない。

これに対して、ランダムユニタリ変換に基づく秘匿計算が提案されている [5] 。この秘匿計算は、準同型暗号やマルチパーティプロトコルと比較して高速な演算が可能であり、さらに、method of optimal direction (MOD) [6] や K-singular value decomposition (K-SVD) [7] 等の広く普及した信号処理アルゴリズムと併用可能である [8]。しかし、データ分析に用いる信号処理アルゴリズムとの併用については、未だ十分な検討がなされていない。

そこで、本稿では、ビッグデータ分析の手法として広範囲な有効性が確認されている LASSO [9] に着目し、ランダムユニタリ変換により秘匿化されたデータに対して、LASSO 解の導出を通して、分析モデルを構築する。LASSO 解の求解には Coordinate descent algorithm (CDA) [10] を利用する。さらに、ランダムユニタリ変換による秘匿化の機能拡張として、分散した各拠点において個別に秘匿化する分散秘匿化を検討する。本稿の貢献は以下の 2 点である。まず、CDA により得られる LASSO 解、即ち、分析モデルは秘匿化の影響を受けず、秘匿化前の原信号に対して構築する分析モデルと等価なモデルが、秘匿化データを用いた場合も構築可能であることを示す。さらに、分散秘匿化に対して、上記 LASSO 解の保全性に関する理論的保証を与える。つまり、各拠点において個別に秘匿化されたデータを集約し、CDA により LASSO 解を求めたとしても、秘匿化の有無によらず、同一の LASSO 解が導出可能であることを示す。

上述のように、LASSO 解の求解に対して分散秘匿化が可能となれば、集約された秘匿化データを用いて、秘匿化前のデータに対する分析モデルと同一の結果を取得出来るようになる。つまり、データの機密性は確保した上で、大規模なデータを利用した分析が可能となり、分散取得されたきたプライバシー保護が必要なデータに対しても、分析精度の向上が実現される。

2. 秘匿信号領域におけるスパースモデリング

2.1 問題の定式化

観測ベクトル $\boldsymbol{y} = (y_0, \cdots, y_{n-1})^T \in \mathbb{R}^n$ を p 本の特徴ベクトル $\boldsymbol{x}_j = (x_{0,j}, \cdots, x_{n-1,j})^T \in \mathbb{R}^n$, $(j = 0, \cdots, p-1)$ の線形和で表現することを考える。特徴ベクトル $\boldsymbol{x}_j$ の重み係数を w_i とし、重み係数ベクトルを $\boldsymbol{w} = (w_0, \cdots, w_{p-1})^T \in \mathbb{R}^p$ とすると、LASSO [9] と呼ばれる定式化では、重み係数ベクトルを次式の制約条件付き最小化問題の解として求解する。

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2 \quad \text{subject to} \quad ||\boldsymbol{w}||_1 \leq \theta \tag{1}$$

ここで、$\boldsymbol{X}$ は $\boldsymbol{x}_j$ を第 j 列とする行列 $\boldsymbol{X} = (\boldsymbol{x}_0, \cdots, \boldsymbol{x}_{p-1}) \in \mathbb{R}^{n\times p}$ である。θ は調整パラメータと呼ばれ、解のスパース性を調整する役割を果たす。上記の制約条件付き最小化問題はラグランジュの未定乗数法を用いて、以下の最小化問題として定式化できる。

$$\begin{aligned}
&\min_{\boldsymbol{w}\in\mathbb{R}^p} L(\boldsymbol{w}),\\
&L(\boldsymbol{w}) \triangleq \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2 + \lambda||\boldsymbol{w}||_1\\
&\qquad = \frac{1}{2}\sum_{i=0}^{n-1}\left(y_i - \sum_{j=0}^{p-1} x_{i,j}w_j\right)^2 + \lambda\sum_{j=0}^{p-1}|w_j|
\end{aligned} \tag{2}$$

ここで、λ はラグランジュ未定乗数であり、調整パラメータ θ に対して定まるパラメータである。なお、以下の議論では、$\sum_{j=0}^{p-1} x_{i,j} = 0$, $\sum_{j=0}^{p-1} x_{i,j}^2 = 1$ であることを仮定する。

2.2 Coordinate descent algorithm

coordinate descent algorithm (CDA) [10] では、式 (2) において、制御対象を 1 つの変数に限定し、残りの変数は変化させない方策をとる。制御対象を w_d とした場合、w_d を含まない項と w_d を含む項に分けるて表記することを考える。ここで、

$$y_i - \sum_{j=0}^{p-1} x_{i,j}w_j = y_i - \sum_{j\neq d}^{p-1} x_{i,j}w_j - x_{i,d}w_d$$

$$\lambda||\boldsymbol{w}||_1 = \lambda|w_d| + \lambda\sum_{j\neq d}^{p-1}|w_j|$$

となることから、

$$r_{i,d} \triangleq y_i - \sum_{j\neq d}^{p-1} x_{i,j}w_j \tag{3}$$

$$c_d \triangleq \lambda\sum_{j\neq d}^{p-1}|w_j|$$

とおくと、式 (2) の評価関数は次式のよう表せる。

$$L(\boldsymbol{w}) = L(w_d, \boldsymbol{w}_{\neq d}) = \frac{1}{2}\sum_{i=0}^{n-1}(r_{i,d} - w_d x_{i,d})^2 + \lambda|w_d| + c_d$$

(4)

ここで、$\boldsymbol{w}_{\neq d} = (w_j | j = 0, \cdots, p-1, j \neq d)^T$ である。上記 $L(w_d, \boldsymbol{w}_{\neq d})$ の最小化の制御対象を w_d に固定した場合、

$$\min_{w_d \in \mathbb{R}} L(w_d, \boldsymbol{w}_{\neq d}) \tag{5}$$

となり、上式の最適解は次式として求まる。

$$w_d^* = \frac{1}{||\boldsymbol{x}_d||_2^2} \text{sign}\left(\boldsymbol{r}_d^T \boldsymbol{x}_d\right) \left(\boldsymbol{r}_d^T \boldsymbol{x}_d - \lambda\right)_+ \tag{6}$$

ここで、$\boldsymbol{r}_d = (r_{0,d}, \cdots, r_{n-1,d})^T$, $\boldsymbol{x}_d = (x_{0,d}, \cdots, x_{n-1,d})^T$ であり、sign$(\cdot)$ および $(\cdot)_+$ は次式の通りである。

$$\text{sign}(t) = \begin{cases} +1 & (t > 0) \\ -1 & (\text{otherwise})o \end{cases}$$

$$(t)_+ = \begin{cases} t & (t > 0) \\ 0 & (\text{otherwise}) \end{cases}$$

同様の処理を $d = 0, 1, \cdots, p-1, 0, 1, \cdots$ として、反復終了条件を満たすまで繰り返す。

2.3 Secure coordinate descent algorithm

ランダムユニタリ行列 $\boldsymbol{Q}_p$ による変換 (ランダムユニタリ変換) を施された信号に対して、CDA により、式 (2) の解を求める。以下での議論の準備として、各信号のランダムユニタリ変換を次のように定義する。

$$\hat{\boldsymbol{y}} \triangleq \boldsymbol{Q}_p \boldsymbol{y} \tag{7}$$

$$\hat{\boldsymbol{X}} \triangleq \boldsymbol{Q}_p \boldsymbol{X} \tag{8}$$

$$\hat{\boldsymbol{x}}_j \triangleq \boldsymbol{Q}_p \boldsymbol{x}_j \tag{9}$$

なお、このとき、

$$\begin{aligned} \hat{\boldsymbol{X}} &= \boldsymbol{Q}_p \boldsymbol{X} \\ &= \boldsymbol{Q}_p (\boldsymbol{x}_0, \cdots, \boldsymbol{x}_p) \\ &= (\boldsymbol{Q}_p \boldsymbol{x}_0, \cdots, \boldsymbol{Q}_p \boldsymbol{x}_p) \\ &= (\hat{\boldsymbol{x}}_0, \cdots, \hat{\boldsymbol{x}}_p) \end{aligned}$$

の関係にあることに注意する。

ここで、解くべきは次式の最小化問題である。

$$\min_{\hat{w}_d \in \mathbb{R}} L(\hat{w}_d, \hat{\boldsymbol{w}}_{\neq d}) \tag{10}$$

$$L(\hat{w}_d, \hat{\boldsymbol{w}}_{\neq d}) = \frac{1}{2} \sum_{i=0}^{n-1} (\hat{r}_{i,d} - \hat{w}_d \hat{x}_{i,d})^2 + \lambda |\hat{w}_d| + \hat{c}_d \tag{11}$$

ここで、

$$\hat{r}_{i,d} \triangleq \hat{y}_i - \sum_{j \neq d}^{p-1} \hat{x}_{i,j} \hat{w}_j \tag{12}$$

$$\hat{c}_d \triangleq \lambda \sum_{j \neq d}^{p-1} \hat{w}_j$$

である。式 (10) における $\hat{w}_d$ の最適解は、式 (6) と同様、次式により得られる。

$$\hat{w}_d^* = \frac{1}{||\hat{\boldsymbol{x}}_d||_2^2} \text{sign}\left(\hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d\right) \left(\hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d - \lambda\right)_+ \tag{13}$$

実は、上式で得られた $\hat{w}_d^*$ は、式 (6) で得られた w_d^* と等価となる。この証明を以下に与える。なお、本証明の骨子は、$\hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d$ と $\boldsymbol{r}_d^T \boldsymbol{x}_d$ 等価性、および、$||\hat{\boldsymbol{x}}_d||_2^2$ と $||\boldsymbol{x}_d||_2^2$ の等価性を導出し、さらに、これら構成要素の等価性に基づき式 (13) と式 (6) の等価性を示すことである。$\hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d$ と $\boldsymbol{r}_d^T \boldsymbol{x}_d$ の関係を考えるのに先立ち、$\hat{\boldsymbol{r}}_d$ と $\boldsymbol{r}_d$ の関係を明らかにする。$\hat{\boldsymbol{r}}_d = (\hat{r}_0, \cdots, \hat{r}_{p-1})^T$ は、式 (12) の定義より、次式のように表すことが出来る。

$$\hat{\boldsymbol{r}}_d = \hat{\boldsymbol{y}} - \sum_{j \neq d}^{p-1} \hat{w}_j \hat{\boldsymbol{x}}_j$$

上式の形式に着目すると、式 (7)(9) より、次のように、$\boldsymbol{r}_d$ との関係を得ることが出来る。

$$\begin{aligned} \hat{\boldsymbol{r}}_d &= \hat{\boldsymbol{y}} - \sum_{j \neq d}^{p-1} \hat{w}_j \hat{\boldsymbol{x}}_j \\ &= \boldsymbol{Q}_p \boldsymbol{y} - \sum_{j \neq d}^{p-1} \hat{w}_j \boldsymbol{Q}_p \boldsymbol{x}_j \\ &= \boldsymbol{Q}_p \left(\boldsymbol{y} - \sum_{j \neq d}^{p-1} \hat{w}_j \boldsymbol{x}_j \right) \\ &= \boldsymbol{Q}_p \boldsymbol{r}_d \end{aligned}$$

上式および式 (9) により、$\hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d$ と $\boldsymbol{r}_d^T \boldsymbol{x}_d$ は等価性は、以下の通り示される。

$$\begin{aligned} \hat{\boldsymbol{r}}_d^T \hat{\boldsymbol{x}}_d &= (\boldsymbol{Q}_p \boldsymbol{r}_d)^T \boldsymbol{Q}_p \boldsymbol{x}_d \\ &= \boldsymbol{r}_d^T \boldsymbol{Q}_p^T \boldsymbol{Q}_p \boldsymbol{x}_d \\ &= \boldsymbol{r}_d^T \boldsymbol{x}_d \end{aligned} \tag{14}$$

$||\hat{\boldsymbol{x}}_d||_2^2$ と $||\boldsymbol{x}_d||_2^2$ の等価性は、式 (9) により以下の通り示される。

$$\begin{aligned} ||\hat{\boldsymbol{x}}_d||_2^2 &= (\boldsymbol{Q}_p \boldsymbol{x}_d)^T \boldsymbol{Q}_p \boldsymbol{x}_d \\ &= \boldsymbol{x}_d^T \boldsymbol{Q}_p^T \boldsymbol{Q}_p \boldsymbol{x}_d \\ &= ||\boldsymbol{x}_d||_2^2 \end{aligned} \tag{15}$$

式 (15)(14) を式 (13) に代入し、式 (6) と比較することで、次式の関係が得られる。

$$\hat{w}_d^* = w_d^* \quad \text{(Q.E.D)}$$

つまり、CD により得られる LASSO 解はランダムユニタリ変換の適用後も不変であることが明らかとなった。

3. 分散秘匿化

複数の拠点で個別にランダムユニタリ変換される場合を考える。拠点数を K とし、第 k 拠点 $(k = 0, \cdots, K-1)$ において取得される観測ベクトルおよび特徴ベクトルを $\boldsymbol{y}^{(k)} \in \mathbb{R}^{n_k}$, $\boldsymbol{X}^{(k)} \in \mathbb{R}^{n_k \times p_k}$ とし、ランダムユニタリ変換に用いるランダ

ムユニタリ行列を $\boldsymbol{Q}_p^{(k)} \in \mathbb{R}^{n_k \times n_k}$ とする。このとき、第 k 拠点において、秘匿化により以下の信号を得る。

$$\hat{\boldsymbol{y}}^{(k)} = \boldsymbol{Q}_p^{(k)} \boldsymbol{y}^{(k)} \tag{16}$$

$$\hat{\boldsymbol{X}}^{(k)} = \boldsymbol{Q}_p^{(k)} \boldsymbol{X}^{(k)} \tag{17}$$

次に、各拠点で秘匿化された信号を集約する。集約した信号を分析するため、$\hat{\boldsymbol{y}}^{(k)}$ および $\hat{\boldsymbol{X}}^{(k)}$ を各々、k に対して昇順に、行方向に連結したベクトル $\hat{\boldsymbol{y}}^{(0:K-1)} \in \mathbb{R}^{Kn_k}$, および行列 $\hat{\boldsymbol{X}}^{(0:K-1)} \in \mathbb{R}^{Kn_k \times n_k}$ を得る。$\hat{\boldsymbol{y}}^{(0:K-1)}$ の第 $kn_k + j$ 要素は、$\hat{\boldsymbol{y}}^{(k)}$ の第 j 要素である。$\hat{\boldsymbol{X}}^{(0:K-1)}$ の第 $kn_k + j$ 行は、$\hat{\boldsymbol{X}}^{(k)}$ のの第 j 行ベクトルである。以下の議論のため、$\boldsymbol{y}^{(k)}$ および $\boldsymbol{X}^{(k)}$ を k に対して昇順に、行方向に連結したベクトルおよび行列を各々、$\boldsymbol{y}^{(0:K-1)} \in \mathbb{R}^{Kn_k}$, $\boldsymbol{X}^{(0:K-1)} \in \mathbb{R}^{Kn_k \times n_k}$ とする。

$\hat{\boldsymbol{y}}^{(0:K-1)}$ と $\boldsymbol{y}^{(0:K-1)}$ の関係、および、$\hat{\boldsymbol{X}}^{(0:K-1)}$ と $\boldsymbol{X}^{(0:K-1)}$ の関係を整理する。

$$\hat{\boldsymbol{y}}^{(0:K-1)} = \boldsymbol{Q}_p^{(0:K-1)} \boldsymbol{y}^{(0:K-1)} \tag{18}$$

$$\hat{\boldsymbol{X}}^{(0:K-1)} = \boldsymbol{Q}_p^{(0:K-1)} \boldsymbol{X}^{(0:K-1)} \tag{19}$$

このとき、$\boldsymbol{Q}_p^{(0:K-1)}$ は、次式の通り、ブロック対角化行列として構成される。

$$\boldsymbol{Q}_p^{(0:K-1)} = \begin{pmatrix} \boldsymbol{Q}_p^{(0)} & & & \boldsymbol{0} \\ & \ddots & & \\ & & \boldsymbol{Q}_p^{(k)} & \\ & & & \ddots \\ \boldsymbol{0} & & & \boldsymbol{Q}_p^{(K-1)} \end{pmatrix} \tag{20}$$

このとき、各ブロック対角要素行列である $\boldsymbol{Q}_p^{(k)}$ がユニタリ行列であることから、

$$(\boldsymbol{Q}_p^{(0:K-1)})^T \boldsymbol{Q}_p^{(0:K-1)} = \boldsymbol{Q}_p^{(0:K-1)} (\boldsymbol{Q}_p^{(0:K-1)})^T$$

は単位行列となり、$\boldsymbol{Q}_p^{(0:K-1)}$ はユニタリ行列であることが分かる。

このように、上述の集約によって得られる秘匿化された信号は、秘匿化前の原信号のランダムユニタリ変換であると言える。従って、ランダムユニタリ変換に対して成立する LASSO 解の保全性は、分散秘匿化に対しても成り立つことが分かる。

4. 実　　験

分散秘匿化されたデータを集約して分析する効果を検証するために、医療分野の臨床データ解析として、糖尿病データを用いて以下のような実験を行った。用いた糖尿病データ [11] [12] は、442 人の患者のデータから構成され、患者に対して 10 項目の検査結果と検査から 1 年後の疾病進行度をデータとして含む。同データに対する分析として、10 項目の検査結果から疾病進行度を予測する予測モデルを構築し、疾病進行度の予測精度を検証対象とした。また、擬似的な分散秘匿化を行うため、上記糖尿病データを K 個のサブセットに分割し、各サブセットが拠点で観測されるデータとみなした。なお、拠点数は $K = 2, 4, 8, 16$ とし、各サブセット内のデータを学習データと検証データに分離した。その上で、以下の 2 種類の予測モデルを比較した。一つ目の予測モデルは、分散秘匿化を用いて全拠点内の学習データを用いて構築した。まず、第 k $(k = 0, \cdots, K-1)$ 拠点において、ランダムユニタリ行列 $\boldsymbol{Q}_{p_k}$ により、学習データ内の検査結果 $\boldsymbol{X}^{(k)}$ および疾病進行度 $\boldsymbol{y}^{(k)}$ を秘匿信号 $\hat{\boldsymbol{y}}^{(k)}$ および $\hat{\boldsymbol{X}}^{(k)}$ に変換した。次に、全拠点内の学習データを秘匿信号として集約し、SCDA により予測モデルを構築した。最後に、同予測モデルを用いて、各拠点の検証データに対して、予測を実施した。以下では、上記予測モデルを統合予測モデルと呼ぶ。二つ目の予測モデルは、自拠点内の学習データのみを用いて構築した。自拠点内の学習データを用いて、CDA により予測モデルを構築し、拠点毎に構築した予測モデルを用いて、各拠点の検証データに対して、予測を実施した。以下では、上記予測モデルを独立予測モデルと呼ぶ。

表 1 に統合予測モデルおよび独立予測モデルにより得られる予測誤差を示す。あわせて、次式の尺度を用いて、統合予測モデルにより達成される予測誤差低減量も評価した。

$$\text{予測誤差低減率} = \frac{\text{独立予測モデルの予測誤差} - \text{統合予測モデルの予測誤差}}{\text{独立予測モデルの予測誤差}}$$

表 1　予 測 誤 差

拠点数	予測誤差: 独立予測モデル	予測誤差: 統合予測モデル	予測誤差 低減率 [%]
4	3236434	3377124	4.2
6	3236434	3286151	1.5
8	3236434	3520810	8.1
10	3236434	3526215	8.2

同表の結果から、統合予測モデルは独立予測モデルに比べて予測誤差を低減できており、各拠点で分散して得られたデータを集約して予測することにより、予測精度の向上に繋がることを確認できた。従来、個人情報等を含むために拠点内に閉じた利用に限定されていたデータであっても、提案技術により、分散取得されたデータを統合した状態での分析が可能となり、分析性能の向上を実現できることを、本実験結果は示している。

さらに、ランダムユニタリ変換による分散秘匿化の秘匿化強度について検証した。ここでは、ある拠点のユーザが、他拠点で秘匿化されたデータにアクセスする場合を想定した。ただし、このユーザは、他拠点のデータの復号に必要なランダムユニタリ行列を知らないため、自拠点のランダムユニタリ行列を用いて秘匿化信号の復号を試みることとした。

5. ま と め

本稿では、分散秘匿化されたデータ上での予測モデル構築について検討した。ランダムユニタリ変換により分散暗号化されたデータに対して、CDA による LASSO 解が秘匿化の前後で保全される事の理論的保証を与えた。さらに、実データを用いた検証を通して、分散秘匿化されたデータを集約して予測モデルを構築することで、予測精度が向上することを検証した。

文　　献

[1] R. L. Lagendijk, Z. Erkin, and M. Barni, "Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation," IEEE Signal Process. Mag., vol.30, no.1, pp.82–105, 2013.

[2] R. Lazzeretti and M. Barni, "Private computing with garbled circuits [applications corner]," IEEE Signal Process. Mag., vol.30, no.2, pp.123–127, 2013.

[3] M. Barni, G. Droandi, and R. Lazzeretti, "Privacy protection in biometric-based recognition systems: A marriage between cryptography and signal processing," IEEE Signal Process. Mag., vol.32, no.5, pp.66–76, 2015.

[4] Z. Brakerski, "Fundamentals of fully homomorphic encryption - A survey," Electronic Colloquium on Computational Complexity, report no. 125, 2018.

[5] I. Nakamura, Y. Tonomura, and H. Kiya, "Unitary transform-based tempalte protection and its application to l2-norm minimization problems," IEICE Trans. Inf. & Syst., vol.E99-D, no.1, p.60 68, 2016.

[6] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp.2443–2446, 1999.

[7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., 2006.

[8] T. Nakachi, Y. Bandoh, and H. Kiya, "Secure overcomplete dictionary learning for sparse representation," IEICE Trans. Inf. & Syst., vol.E103-D, no.1, 2020.

[9] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society, Series B, vol.58, no.1, pp.267–288, 1996.

[10] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," Annals of Applied Statistics, vol.1, no.2, pp.302–332, 2007.

[11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," Annals of Statistics, vol.32, no.2, pp.407–499, 2004.

[12] "`https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html`,".