ランダムユニタリ変換ベースの分散秘匿化信号に対するスパースモデリング

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　〒 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。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。ランダムユニタリ変換に基づく秘匿計算は、マルチパーティプロトコルや準同型暗号と比較して高速な演算が可能であり、さらに、既存の信号処理アルゴリズムと併用可能であるというメリットがある。しかし、分散した拠点毎に個別に秘匿化する分散秘匿化については、十分な検討がなされていない。そこで、本稿では、ランダムユニタリ変換による分散秘匿化に対して、スパースモデリング (Elastic Net) の解の保全性に関する理論的保証を与える。さらに、各拠点において取得可能なサンプル数が少数の場合、ランダムユニタリ変換の秘匿化強度が低下する問題に対して、同変換の次元拡張に基づき、秘匿化強度を向上させる手法についても提案する。本稿で提案する分散秘匿化により、プライバシー保護の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　スパースモデリング、Elastic Net、ランダムユニタリ変換

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 [1] [2] [3] [4]。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の場合、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、医療機関において取得した臨床検査の結果 [5]、スマートシティ [6] や監視サービス [7] 等において取得した画像データのように特定個人を識別できる情報である。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 [8] [9] [10]。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 [11] [12] [13] [14] 。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 [15] 。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、スパース信号表現 [16] [17]、画像圧縮 [18] [19] 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。こうした特性を活かして、秘匿領域における信号処理アルゴリズム（例：秘匿領域におけるスパース表現のための辞書学習 [20]、秘匿領域における画像信号圧縮 [21] 等）が研究されている。しかし、拠点毎に独立に秘匿化する分散秘匿化、および、分散秘匿化されたデータの分析については、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 Elastic Net [22] に着目し、ランダムユニタリ変換により分散秘匿化されたデータに対して、Elastic Net 解の導出を通して、分析モデルを構築する。本稿の貢献は以下の 2 点である。まず、各拠点において個別に秘匿化されたデータを集約し、Elastic Net 解を求めたとしても、秘匿化の有無によらず、同一の解が導出可能であることを示す。つまり、分散秘匿化に対して、上記 Elastic Net 解の保全性に関する理論的保証を与える。さらに、拠点で観測可能なサンプル数が少ない場合、ランダムユニタリ変換の秘匿化強度が低下する問題に対して、次元拡張に基づく秘匿化強度の強化法を提案する。上述のように、Elastic Net 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 秘匿信号に対する分析

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、Elastic Net [22] 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。

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2 \quad \text{s.t. } \alpha||\boldsymbol{w}||_1 + (1-\alpha)||\boldsymbol{w}||_2^2 \leqq \theta \quad (1)$$

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。α は L1 ノルムと L2 ノルムの比率を調整する役割を果たし、θ は制約条件の強さ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。

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

ここで、ρ はパラメータ 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 ランダムユニタリ変換

ランダムユニタリ行列 $\boldsymbol{Q}_\zeta$ による変換 (ランダムユニタリ変換) に基づき秘匿化された信号に対して、式 (2) の解を求めることを考える。ランダムユニタリ変換に基づく秘匿演算では、鍵 p によって生成されるランダムユニタリ行列 $\boldsymbol{Q}_\zeta$ を用いた変換により、対象信号 $\boldsymbol{y}$ を秘匿信号 $\boldsymbol{Q}_\zeta\boldsymbol{y}$ へ変換する。このとき、$\boldsymbol{Q}_\zeta \in \mathbb{R}^{n\times n}$ であり、以下を満たす：

$$\boldsymbol{Q}_\zeta^*\boldsymbol{Q}_\zeta = \boldsymbol{I} \qquad (3)$$

ここで、$[\cdot]^*$ はエルミート転置 [1]、$\boldsymbol{I}$ は単位行列を表す。ランダムユニタリ行列の生成には、複数のユニタリ行列を組み合わせることで $\boldsymbol{Q}_\zeta$ を生成する方法や、擬似乱数行列に対してグラムシュミットの直交化を適用する方法が提案されている [15]。本稿では、後者の方法により生成したランダムユニタリ行列を用いる。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。

$$\hat{\boldsymbol{y}} \triangleq \boldsymbol{Q}_\zeta\boldsymbol{y} \qquad (4)$$

$$\hat{\boldsymbol{X}} \triangleq \boldsymbol{Q}_\zeta\boldsymbol{X} \qquad (5)$$

$$\hat{\boldsymbol{x}}_j \triangleq \boldsymbol{Q}_\zeta\boldsymbol{x}_j \qquad (6)$$

なお、このとき、

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

の関係にあることに注意する。

秘匿化された信号に対して Elastic Net 解を求めるには、以下のコストを最小化する必要がある。

$$\hat{L}(\boldsymbol{w}) \triangleq \frac{1}{2}||\hat{\boldsymbol{y}} - \hat{\boldsymbol{X}}\boldsymbol{w}||_2^2 + \rho\alpha||\boldsymbol{w}||_1 + \rho(1-\alpha)||\boldsymbol{w}||_2^2 \qquad (7)$$

このとき、式 (3) より、次の関係が成り立つことがわかる。

$$\arg\min_{\boldsymbol{w}} \hat{L}(\boldsymbol{w}) = \arg\min_{\boldsymbol{w}} L(\boldsymbol{w})$$

上式は、秘匿化された信号に対して求めた Elastic Net 解は、秘匿化前の信号に対する解と一致することを示している。

3 分散秘匿化

拠点毎にランダムユニタリ行列を設定し、各拠点において独立にデータを秘匿化する分散秘匿化を考える。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 $\boldsymbol{y}^{(k)} \in \mathbb{R}^{n_k}$, $\boldsymbol{X}^{(k)} \in \mathbb{R}^{n_k\times p}$ 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 $\boldsymbol{Q}_\zeta^{(k)} \in \mathbb{R}^{n_k\times n_k}$ 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。

$$\hat{\boldsymbol{y}}^{(k)} = \boldsymbol{Q}_\zeta^{(k)}\boldsymbol{y}^{(k)} \qquad (8)$$

$$\hat{\boldsymbol{X}}^{(k)} = \boldsymbol{Q}_\zeta^{(k)}\boldsymbol{X}^{(k)} \qquad (9)$$

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 $\hat{\boldsymbol{y}}^{(0:K-1)} \in \mathbb{R}^N$, および行列 $\hat{\boldsymbol{X}}^{(0:K-1)} \in \mathbb{R}^{N\times p}$ を得る。ここで、$N = \sum_{k=0}^{K-1} n_k$ とする。$\hat{\boldsymbol{y}}^{(0:K-1)}$ の第 $(j + \sum_{k'=0}^{k-1} n_{k'})$ 要素は、$\hat{\boldsymbol{y}}^{(k)}$ の第 j 要素である。$\hat{\boldsymbol{X}}^{(0:K-1)}$ の第 $(j + \sum_{k'=0}^{k-1} n_{k'})$ 行は、$\hat{\boldsymbol{X}}^{(k)}$ 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、$\boldsymbol{y}^{(0:K-1)} \in \mathbb{R}^N$, $\boldsymbol{X}^{(0:K-1)} \in \mathbb{R}^{N\times p}$ 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 $\boldsymbol{X}^{(0:K-1)}$ の関係を以下の通り整理する。

$$\hat{\boldsymbol{y}}^{(0:K-1)} = \boldsymbol{Q}_\zeta^{(0:K-1)}\boldsymbol{y}^{(0:K-1)} \qquad (10)$$

$$\hat{\boldsymbol{X}}^{(0:K-1)} = \boldsymbol{Q}_\zeta^{(0:K-1)}\boldsymbol{X}^{(0:K-1)} \qquad (11)$$

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

1：本稿では、$\boldsymbol{Q}_\zeta$ として実数要素を持つ行列を考えるため、エルミート転置は単なる転置演算として議論している。

$$\boldsymbol{Q}_{\zeta}^{(0:K-1)} = \begin{pmatrix} \boldsymbol{Q}_{\zeta}^{(0)} & & & \boldsymbol{0} \\ & \ddots & & \\ & & \boldsymbol{Q}_{\zeta}^{(k)} & \\ & & & \ddots \\ \boldsymbol{0} & & & \boldsymbol{Q}_{\zeta}^{(K-1)} \end{pmatrix} \quad (12)$$

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

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

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

このように、たとえ拠点毎に異なるランダムユニタリ行列を用いて独立に秘匿化したとしても、上述の集約によって得られる秘匿化された信号は、秘匿化前の全ての信号に対するランダムユニタリ変換（式 (12) により規定）となっている。従って、ランダムユニタリ変換に対して成立する Elastic Net 解の保全性は、上述の分散秘匿化に対しても成り立つことが分かる。

4 次元拡張による秘匿性強化

$\boldsymbol{y} \in \mathbb{R}^n$ および $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ を各々、$\tilde{n}$ 次元ベクトルおよび $\tilde{n} \times p$ 行列に拡張 $(\tilde{n} > n)$ し、秘匿性の強化を図ることを考える。これは、ランダムユニタリ行列のサイズを大きくすることにより、鍵空間を拡大でき、秘匿性を強化できるためである。

そこで、$\tilde{n} \times \tilde{n}$ サイズのランダムユニタリ行列[2] $\boldsymbol{Q}_{\zeta,\tilde{n}} \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$ を $\boldsymbol{y}$, $\boldsymbol{X}$ の秘匿化に用いることとし、秘匿化後の信号 $\tilde{\boldsymbol{y}} \in \mathbb{R}^{\tilde{n}}$, $\tilde{\boldsymbol{X}} \in \mathbb{R}^{\tilde{n} \times p}$ を次のような変換により得ることを考える。

$$\tilde{\boldsymbol{y}} = \boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S} \boldsymbol{y} + \boldsymbol{\psi} \quad (13)$$

$$\tilde{\boldsymbol{X}} = \boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S} \boldsymbol{X} + \boldsymbol{\Phi} \quad (14)$$

ここで、$\boldsymbol{S} \in \mathbb{R}^{\tilde{n} \times n}$ は、ベクトルの次元を n から $\tilde{n}$ へ拡張する変換である。また、$\boldsymbol{\psi} \in \mathbb{R}^{\tilde{n}}$ および $\boldsymbol{\Phi} \in \mathbb{R}^{\tilde{n} \times p}$ は、次式を満たす $\tilde{n}$ 次元ベクトルおよび $\tilde{n} \times p$ サイズの行列として設定する。

$$\boldsymbol{\psi}^T (\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}) = \boldsymbol{0}_n \in \mathbb{R}^n \quad (15)$$

$$\boldsymbol{\Phi}^T (\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}) = \boldsymbol{0}_{p \times n} \in \mathbb{R}^{p \times n} \quad (16)$$

$$\boldsymbol{\psi}^T \boldsymbol{\Phi} = \boldsymbol{0}_p \in \mathbb{R}^p \quad (17)$$

ここで、$\boldsymbol{0}_n$、$\boldsymbol{0}_p$ および $\boldsymbol{0}_{p \times n}$ は、各々、全ての要素を 0 とする n 次元ベクトル、p 次元ベクトルおよび $p \times n$ 行列である。つまり、上式は、$\boldsymbol{\psi}$ が $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の列ベクトルと直交すること、そして、$\boldsymbol{\Phi}$ の列ベクトルが $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の列ベクトルと直交すること、さらに、$\boldsymbol{\psi}$ が $\boldsymbol{\Phi}$ の列ベクトルと直交することを、各々、要請する。

以下、$\boldsymbol{S}$、$\boldsymbol{\psi}$、および $\boldsymbol{\Phi}$ の設定法について説明する。まず、$\boldsymbol{S}$ の設定法を示す。$\boldsymbol{S} \in \mathbb{R}^{\tilde{n} \times n}$ は、要素として、0,1 のいずれかをとり、各列の要素として 1 をひとつだけ含み、さらに、各列で 1 を取る行を列毎に異なるように設定する。この $\boldsymbol{S}$ を用いた変換 $\boldsymbol{Sy}$ により、$\tilde{n}$ 個の要素のうち n 個が $\boldsymbol{y}$ の要素であり、残りの要素を 0 とする $\tilde{n}$ 次元ベクトルを得る。次に、$\boldsymbol{\psi}$ の設定法を示す。$\boldsymbol{\psi}$ は、$\boldsymbol{Q}_{\zeta,\tilde{n}}$ の列ベクトルを用いて構成する。$\boldsymbol{S}$ の i 行 j 列 $(i = 1, \cdots, \tilde{n}, j = 1, \cdots, n)$ の要素を $s_{i,j}$ として、$\boldsymbol{S}$ の j 列において要素を 1 とする行のインデックスを $\tilde{i}(j)$ で表すものとする。この場合、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ は、n 本の列ベクトルからなる行列であり、その n 本の列ベクトルは、$\boldsymbol{Q}_{\zeta,\tilde{n}}$ の第 $\tilde{i}(j)$ 列ベクトル $(j = 1, \cdots, n)$ である。$\boldsymbol{Q}_{\zeta,\tilde{n}}$ の列ベクトルのうち、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ に含まれない列ベクトルは、$\tilde{n} - n$ 本存在する。この $\tilde{n} - n$ 本の列ベクトルを $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルと呼ぶことにする。$\boldsymbol{Q}_{\zeta,\tilde{n}}$ がユニタリ行列であることから、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルは $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の列ベクトルと直交する。そこで、$\boldsymbol{\psi}$ として、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルのいずれか一つを設定することで、条件 (式 (15)) を満たす形で $\boldsymbol{\psi}$ を構成できる。最後に、$\boldsymbol{\Phi}$ の設定法を示す。$\boldsymbol{\psi}$ として選択されていない $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルから、p 本の列ベクトルを選択し、これらを列ベクトルとする行列として、$\boldsymbol{\Phi}$ を構成する。前述と同様の理由により、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルが $\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の列ベクトルと直交することから、$\boldsymbol{\Phi}$ は条件 (式 (16)) を満たす形で構成できる。また、$\boldsymbol{Q}_{\zeta,\tilde{n}} \boldsymbol{S}$ の直交補空間ベクトルの列ベクトル同士も直交することから、条件 (式 (17)) も満たすことができる。

ここで、次のようなコスト関数を定義する。

$$\tilde{L}(\boldsymbol{w}) \triangleq \frac{1}{2} ||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}} \boldsymbol{w}||_2^2 + \rho\alpha ||\boldsymbol{w}||_1 + (\rho(1-\alpha) - \frac{1}{2}) ||\boldsymbol{w}||_2^2 \quad (18)$$

このとき、

$$||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}} \boldsymbol{w}||_2^2 = ||\boldsymbol{y} - \boldsymbol{X} \boldsymbol{w}||_2^2 + ||\boldsymbol{w}||_2^2 + ||\boldsymbol{\psi}||_2^2 \quad (19)$$

であることから (証明は付録 **1** を参照)、

$$\begin{aligned} \tilde{L}(\boldsymbol{w}) &= ||\boldsymbol{y} - \boldsymbol{X} \boldsymbol{w}||_2^2 + \rho\alpha ||\boldsymbol{w}||_1 + \rho(1-\alpha) ||\boldsymbol{w}||_2^2 + \frac{1}{2} ||\boldsymbol{\psi}||_2^2 \\ &= L(\boldsymbol{w}) + \frac{1}{2} ||\boldsymbol{\psi}||_2^2 \end{aligned}$$

となる。従って、次の関係が成り立つことがわかる。

$$\arg\min_{\boldsymbol{w}} \tilde{L}(\boldsymbol{w}) = \arg\min_{\boldsymbol{w}} L(\boldsymbol{w})$$

つまり、式 (13)(14) により秘匿化された信号に対して、$\arg\min_{\boldsymbol{w}} \tilde{L}(\boldsymbol{w})$ を最小化する解を求めれば、秘匿化前の信号に対する Elastic Net 解を導出できることを上式は示している。

5 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 [23] [24] 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

2 : $\boldsymbol{Q}_{\zeta,\tilde{n}} \in \mathbb{R}^{\tilde{n} \times \tilde{n}}$ は、行列のサイズを陽に示す表記に変更している。

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 $\boldsymbol{Q}_{\zeta_k}$ 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 $\hat{\boldsymbol{X}}^{(k)}$ および $\hat{\boldsymbol{y}}^{(k)}$ 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、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、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。

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

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。また、同表によれば、拠点数の増加に伴い、独立予測モデルの予測誤差が増加し、統合予測モデルによる予測誤差低減率が向上していることが確認できる。これは、本実験における拠点当たりの学習データ数は、拠点数に反比例して少なくなるように設定しているためと考える。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。秘匿化前の原信号と上記の手順で復号された信号の類似度を両信号の相関係数により評価した。表 2 に、上記相関係数の絶対値の平均値を示す。同表 (a) は、**4** 節に示す方法にて、次元拡張による秘匿性強化を実施した結果である。同表 (b) は上記秘匿性強化を行わない場合の結果である。同表 (a)(b) を比較した結果、他拠点のランダムユニタリ行列により復号される情報が大幅に低減できており、拠点数が 20 の場合のように、拠点当たりの学習データ数が少ない場合でもあっても、十分に安全と考えられる程度 (相関係数が 1 未満) まで、秘匿化されていることが確認できる。これにより、観測されるサンプル数が少ない場合（つまり、$\hat{\boldsymbol{y}}^{(k)}$ の要素数および $\hat{\boldsymbol{X}}^{(k)}$ の行数が小さな値の場合）、ランダムユニタリ変換の鍵空間が小さくなったとしても、提案法により、情報の漏洩を回避するに十分な秘匿性を提供できたことを示している。これにより、閲覧権のないユーザでは秘匿化信号から原信号を復号できるリスクを大幅に低減することが可能となる。

表 1　予 測 誤 差

(a) $\rho = 1.0$ の場合

拠点数	予測誤差: 統合予測モデル	予測誤差: 独立予測モデル	予測誤差 低減率 [%]
4	3236434	3377124	4.2
8	3236434	3520810	8.1
16	3236434	3619758	10.6
20	3236434	3834318	15.6

(b) $\rho = 0.9$ の場合

拠点数	予測誤差: 統合予測モデル	予測誤差: 独立予測モデル	予測誤差 低減率 [%]
4	3271445	3308948	1.1
8	3271445	3378308	3.2
16	3271445	3562837	8.2
20	3271445	3709451	11.8

表 2　他拠点のランダムユニタリ行列による復号復号と原信号の類似度（相関係数の絶対値）

(a) 次元拡張による秘匿強化有

拠点数	4	8	16	20
類似度（検査結果）	0.011	0.017	0.020	0.020
類似度（疾病進行度）	0.026	0.040	0.070	0.092

(b) 次元拡張による秘匿強化無

拠点数	4	8	16	20
類似度（検査結果）	0.049	0.059	0.119	0.112
類似度（疾病進行度）	0.106	0.107	0.165	0.224

6 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る分散秘匿化の前後で Elastic Net 解が保全される事の理論的保証を与え、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実証した。さらに、観測されるサンプル数が少ない場合に、ランダムユニタリ変換の秘匿化強度が低下する問題に対して、次元拡張に基づく手法により、ランダムユニタリ変換の秘匿化強度が向上を実現した。

文　献

[1] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016.

[2] W. Shi and S. Dustdar, “The promise of edge computing,” *Computer*, vol. 49, no. 5, pp. 78–81, 2016.

[3] M. Satyanarayanan, “The emergence of edge computing,” *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[4] G. Premsankar, M. Di Francesco, and T. Taleb, “Edge com-

puting for the internet of things: A case study," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1275–1284, 2018.
[5] H. A. Al Hamid, S. M. M. Rahman, M. S. Hossain, A. Almogren, and A. Alamri, "A security model for preserving the privacy of medical big data in a healthcare cloud using a fog computing facility with pairing-based cryptography," *IEEE Access*, vol. 5, pp. 22313–22328, 2017.
[6] T. Taleb, S. Dutta, A. Ksentini, M. Iqbal, and H. Flinck, "Mobile edge computing potential in making cities smarter," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 38–43, 2017.
[7] R. Xu, S. Y. Nikouei, Y. Chen, A. Polunchenko, S. Song, C. Deng, and T. R. Faughnan, "Real-time human objects tracking for smart surveillance at the edge," *Proc. IEEE Int. Conf. Commun.*, pp. 1–6, 2018.
[8] C. T. Huang, L. Huang, Z. Qin, H. Yuan, L. V. Varadharajan, and C. C. J. Kuo, "Survey on securing data storage in the cloud," *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. e4, 2014.
[9] P. J. Sun, "Privacy protection and data security in cloud computing: A survey, challenges, and solutions," *IEEE Access*, vol. 7, pp. 147420–147452, 2019.
[10] Y. Xiao, Y. Jia, C. Liu, X. Cheng, J. Yu, and W. Lv, "Edge computing security: State of the art and challenges," *Proc. IEEE*, vol. 107, no. 8, pp. 1608–1631, 2019.
[11] 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.
[12] R. Lazzeretti and M. Barni, "Private computing with garbled circuits [applications corner]," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 123–127, 2013.
[13] 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.
[14] Z. Brakerski, "Fundamentals of fully homomorphic encryption - A survey," *Electronic Colloquium on Computational Complexity*, report no. 125, 2018.
[15] 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.
[16] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, pp. 4311–4322, 2006.
[17] M. Elad, "Sparse and redundant representation modeling - what next?," *IEEE Trans. Signal Process. Lett.*, vol. 19, no. 12, pp. 922–928, 2012.
[18] ITU-T and ISO/IEC JTC 1, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines, ITU-T Rec. T.81 and ISO/IEC ISO/IEC 10918-1:1994*, June 1994.
[19] ITU-T and ISO/IEC JTC 1, *Information technology JPEG 2000 image coding system: Core coding system, ITU-T Rec.T.800 and ISO/IEC 15444-1:2004, Edition 3.0*, June 2019.
[20] T. Nakachi, Y. Bandoh, and H. Kiya, "Secure overcomplete dictionary learning for sparse representation," *IEICE Trans. Inf. & Syst.*, vol. E103-D, no. 1, pp. 50–58, 2020.
[21] T. Chuman, K. Iida, W. Sirichotedumrong, and H. Kiya, "Encryption-then-compression systems using grayscale-based image encryption for JPEG images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 6, pp. 1515–1525, 2019.
[22] H.Zou and T.Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
[23] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
[24] *https://www4.stat.ncsu.edu/~boos/var.select/diabetes.html*.

付　　録

1　式 (19) の証明

式 (19) の左辺を以下の通り、展開する。

$$\begin{aligned}
&||\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{w}||_2^2 \\
&= ||\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) + \boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w}||_2^2 \\
&= ||\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})||_2^2 + ||\boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w}||_2^2 \\
&+ 2(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}))^T(\boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w})
\end{aligned}$$

ここで、上式の各項について考える。

まず、第一項については、以下の関係を得る。

$$\begin{aligned}
&||\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})||_2^2 \\
&= (\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}))^T(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})) \\
&= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})^T(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) \\
&= ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}||_2^2
\end{aligned}$$

最後の式展開では、以下の関係を用いた。

$$\begin{aligned}
(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})^T(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}) &= \boldsymbol{S}^T\boldsymbol{Q}_{\zeta,\tilde{n}}^T\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S} \\
&= \boldsymbol{S}^T\boldsymbol{S} \\
&= \boldsymbol{I}_n
\end{aligned}$$

ここで、$\boldsymbol{I}_n$ は $n \times n$ サイズの単位行列である。

次に、第二項については、以下の関係を得る。

$$\begin{aligned}
&||\boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w}||_2^2 \\
&= ||\boldsymbol{\psi}||_2^2 + \boldsymbol{\psi}^T(\boldsymbol{\Phi}\boldsymbol{w}) + (\boldsymbol{\Phi}\boldsymbol{w})^T\boldsymbol{\psi} + ||\boldsymbol{\Phi}\boldsymbol{w}||_2^2 \\
&= ||\boldsymbol{\psi}||_2^2 + ||\boldsymbol{w}||_2^2
\end{aligned}$$

ここで、最後の式展開では、式 (17) および、以下の関係を用いた。

$$\begin{aligned}
||\boldsymbol{\Phi}\boldsymbol{w}||_2^2 &= ||\sum_{i=1}^{p} w_i\boldsymbol{\phi}_i||_2^2 \\
&= (\sum_{j=1}^{p} w_j\boldsymbol{\phi}_j)^T(\sum_{i=1}^{p} w_i\boldsymbol{\phi}_i) \\
&= \sum_{i=1}^{p} w_i^2
\end{aligned}$$

最後に、第三項については、以下の関係を得る。

$$\begin{aligned}
&(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}))^T(\boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w}) \\
&= ((\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}))^T(\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})^T(\boldsymbol{\psi} - \boldsymbol{\Phi}\boldsymbol{w}) \\
&= ((\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}))^T((\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})^T\boldsymbol{\psi} - (\boldsymbol{Q}_{\zeta,\tilde{n}}\boldsymbol{S})^T\boldsymbol{\Phi}\boldsymbol{w}) \\
&= 0
\end{aligned}$$

最後の式展開では、式 (15)(16) の関係を用いた。

この結果、式 (19) の右辺を得る。