教育データ分析への差分プライバシ適用に関する有用性の検討

清水　將吾

Abstract

As the use of ICT in the field of education is advancing, learning analytics, which aims to analyze educational data to improve learning effectiveness, is becoming popular. On the other hand, since learning data is related to privacy issues, its utilization must be carefully considered. In this paper, we focus on differential privacy as a technical privacy protection measure and investigate the usefulness of applying differential privacy to educational data analysis using public datasets.

1　はじめに

LMS やオンライン教育，MOOC など教育分野での ICT 活用が進み，広範囲にわたり学習データが蓄積されるようになっている．これらのデータを分析して教育効果の改善につなげることを目的とするラーニングアナリティスクス（learning analytics; LA）や教育データマイニング（educational data mining; EDM）と呼ばれる分野では，学習者の学習行動の予測や教材や問題の推薦，ダッシュボードの開発といった研究が活発に行われている [1].

一方，学習データはプライバシに深く関わるため，利活用には十分な注意が必要である．一般には成績分布などの統計量から個人の情報を推測することは難しいが，場合によっては特定の個人の情報を高い精度で推測できることがある．簡単な例として，表 1 のようにある科目の合否人数を学科別と学年別の両方で集計する場合を考える．さらに，学科 A でもなく，学年 x でもない学生がいないことは分かっているとする．このとき，学科 A で，かつ学年 x である学生は 3 名いるが，3 名とも否であることが分かる．

このような状況を防ぐためのプライバシ保護技術として，以下が挙げられる [2].

- k-匿名性: 個人を識別できる属性の組について，同じ値をもつレコードが少なくとも k 個以上存在するように値の汎化を行う．例えば，学生の所属と入学時の入試区分を間接的な識別属性とする場合，定員の少ない入試区分では個人が識別されやすくなるため，入試区分の情報の粒度を上げることによって安全性を向上させる．

表 1　プライバシ開示の例

全体	人数
合	15
否	5

学科 A	人数
合	8
否	3

学年 x	人数
合	7
否	5

- 差分プライバシ: 公開される統計量に確率的なノイズを加えることで，統計量から個々のレコードが決定的に推測されるリスクを低減する．
- 秘密計算: 暗号技術を用いて，自身が持っている情報を解析者に開示することなく，計算結果のみを得る．計算過程における入力データの保護を目的とする．

k-匿名性は攻撃者の背景知識や攻撃能力を制限しており，それを超える攻撃が現実に考えられる場合は脆弱になる．一方，差分プライバシは，攻撃者の背景知識や攻撃能力を限定せず，任意の攻撃に対して弱い秘匿性を保証する．

本稿では，差分プライバシを取り上げる．差分プライバシはメカニズムと呼ばれるランダム化手法を通じて，元データにノイズを加えることによってプライバシ保護を実現する．このため，ノイズの量が多いほどプライバシ保護の程度は高まるが，得られるデータと元データとの差が大きくなり，有用性が低くなるというトレードオフの関係がある．出力誤差については理論的な上界が示されているが，実際に確保できる有用性の程度は大きさや分布などのデータの特性による．本稿では，公開されている教育データセットを用いて，差分プライバシを適用してどの程度のプライバシ保護と有用性を達成できるのか検証を行う．

なお，本稿では教育データの利活用に関して技術的な側面のみに焦点をあて，制度的な側面については別途検討するものとする．

2　準備

2.1　教育データ分析

平均や分散などの記述統計，層別分析，可視化は一般的に最初に行われる分析手法である．教育データ分析での基本的な分析手法としては，2 つの試験結果を比較するための t 検定，グループ間で試験結果を比較するための分散分析，指導方法の違いによる影響を確認するための効果量の算出などが挙げられる．また，授業評価などの自由記述に対するテキストマイニングを用いた分析も最近では一般的である．

よりデータ駆動型のアプローチとして LA や EDM でよく使われている分析手法と適用事例は文献 [16, 17] にまとめられている．決定木やニューラルネットワークなどの機械学習を用いた成績予測，学習者の知識状態を推定する知識トレーシング，個別最適化を実現するための教材推薦，評価の自動化などが代表的なタスクである．予測モデルは入力である説明変数の値から出力である目的変数の値を予測するものである．予測モデルの構造に応じたパラメータを，あらかじめ用意された入力と出力の組からなるデータを用いて学習する．特に近年注目を集めている深層学習の本領域への適用事例は文献 [12] で報告されている．教員は得られた分析結果に基づいて適切にフィードバックを与えることが期待される．

分析対象となるデータは年齢，性別，居住地域などのデモグラフィック属性の他に，試験の点数，各問題に対する回答，アンケートの回答として入手できるリッカート尺度や自然言語，ICT ログとして入手できる操作履歴や操作時間，教員の指導履歴，学生間のソーシャルネットワークなど様々存在する．1 授業単位の情報であれば教員個人で収集可能なこともあるが，それを超える規模や種類のデータを必要とする場合は分析手法を検討するためのデータを事前に入手することが困難である．このため，手法の有効性を容易に検証または比較できることを目的として，研究機関や学会，Kaggle のようなクラウドソーシング型のデータ分析プラットフォームから一般に入手可能な大規模教育データセットが提供されている [14].

2.2 プライバシ保護技術

2.2.1 差分プライバシ

差分プライバシは，Dwork [9] によって 2006 年に提案された，統計量に確率的なノイズを加えることによってデータ提供者のプライバシーを保護する技術である．例えば，教員がある科目の成績分布を公開する際に，攻撃者にどのような背景知識があったとしても個人の成績が秘匿されることを数学的に保証したい場合に適用できる．

以下，定義を与える．データベースはレコードの集合である．2 つのデータベース D_1, D_2 が 1 つのレコードだけ異なるとき，D_1 と D_2 は隣接していると言う．ランダム化メカニズム $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ が任意の隣接するデータベース $D_1, D_2 \in \mathcal{D}$ に対して以下の式を満たすとき，$\mathcal{M}$ は ϵ-差分プライバシを満たすと言う．S は $\mathcal{R}$ の任意の部分集合である．

$$\Pr[\mathcal{M}(D_1) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(D_2) \in S]$$

あるデータ提供者 A に着目し，D_1 を A のレコードを含むデータベース，D_2 を A の

レコードを含まない，D_1 と隣接するデータベースとする．差分プライバシの定義は，任意の $s \in \mathcal{R}$ に対して $\Pr[\mathcal{M}(D_1) = s]$ が $\Pr[\mathcal{M}(D_2) = s]$ の e^{ϵ} 倍を超えないことを示している．したがって，ϵ が十分小さければ，攻撃者が s を知ったとしてもそれが D_1 から生成されたものか D_2 から生成されたものか識別できず，したがって A のプライバシを秘匿できる．

ϵ は安全性を調整するパラメータであり，有用性とトレードオフの関係にある．$\epsilon = 0$ のとき最も高い安全性を提供できるが，これを達成できるランダム化が必要となり，有用性は得られない．$\epsilon = \infty$ のとき有用性は最も高いが，ランダム化されないために安全性は得られない．実用では，安全性と有用性のバランスを取りながら ϵ の値を決定する．

本稿では，$\mathcal{M}$ として Laplace メカニズムを採用する．Laplace メカニズムは，出力に Laplace 分布に従った乱数を加えることによって ϵ-差分プライバシを達成する．Laplace メカニズムを用いた場合の有用性について，以下のことが示されている．問合せ q の感度 $\Delta_{1,q}$ を，1 つのレコードが q の出力に与える影響の最大値とする．例えば，レコード数を n，レコード x のドメインを $\{0,1\}$ としたとき，q が和関数であれば $\Delta_{1,q} = 1$，q が平均関数であれば $\Delta_{1,q} = 1/n$，q が最大値関数であれば $\Delta_{1,q} = 1$ である．このとき，任意の $\delta \in (0,1]$ に対して，

$$\Pr\left(||\mathcal{M}(D) - q(D)||_1 > \frac{\Delta_{1,q}}{\epsilon} \ln \frac{1}{\delta}\right) \leq \delta$$

が成り立つ．この定理はメカニズムの出力の誤差が $\mathcal{O}\left(\frac{\Delta_{1,q}}{\epsilon}\right)$ で抑えられることを意味する．

差分プライバシの重要な性質として，以下の合成定理がある．合成定理は，差分プライバシを満たす複数のメカニズムを合成したときに得られる安全性を保証する．

各 $i \in \{1,k\}$ に対して，$\mathcal{M}_i$ を ϵ_i-差分プライバシを満たすメカニズムとする．このとき，$\{\mathcal{M}_i(D)\}_{i=1}^{k}$ を出力するメカニズム $\mathcal{M}$ は $\sum_{i=1}^{k} \epsilon_i$-差分プライバシを満たす．

これによって，複数のステップからなる統計処理であっても，それぞれに差分プライバシが保証された単純な処理に分割することで全体のプライバシを保証できる．さらに，上記の特別な場合として，$\mathcal{M}_i$ が適用されるレコードの集合が互いに素であるとき，$\mathcal{M}$ は $(\max_i \epsilon_i)$-差分プライバシを満たすことが知られている．これを並列合成定理と呼ぶ．

2.2.2　局所差分プライバシ

局所差分プライバシは，統計量を生成するデータ収集者が信頼できない場合にデータ提供者がランダム化を行ってからデータを送信する技術である．例えば，学生が質問に対し

て回答を送信する際，教員に対してその回答を完全には知られたくない場合に適用できる．教員は統計量を生成できればよいので，個々の学生がどのような回答をしたかには興味がなく，加工されたデータの集合から元の統計量を精度良く推定できれば十分である．ただし，後述するように，現実的な有用性を得るためには MOOC やログデータのようなある程度の規模をもつデータへの適用が適切である．

ランダム化メカニズム $\mathcal{M} : \mathcal{S} \to \mathcal{R}$ が任意の異なる入力 $x_1, x_2 \in \mathcal{S}$ に対して以下の式を満たすとき，$\mathcal{M}$ は ϵ-局所差分プライバシを満たすと言う．S は $\mathcal{R}$ の任意の要素である．

$$\Pr[\mathcal{M}(x_1) \in S] \leq e^{\epsilon} \Pr[\mathcal{M}(x_2) \in S]$$

局所差分プライバシを実現する簡単なアルゴリズムとして，ランダム化応答（randomized response）がある．データ提供者が持っているデータを $x \in \{0, 1\}$ とする．ランダム化応答の結果 z は次式で与えられる．

$$z = \begin{cases} x, & \text{with probability } e^{\epsilon}/(1+e^{\epsilon}) \\ 1-x, & \text{with probability } 1/(1+e^{\epsilon}) \end{cases}$$

上の式は，確率 $1/(1+e^{\epsilon})$ で x を反転して送信することを表している．x をそのまま送信する確率が x を反転して送信する確率の e^{ϵ} 倍になっており，ϵ-局所差分プライバシを満たすことが分かる．この定義では $\epsilon = 0$ のとき確率 50% で，$\epsilon = 5$ では確率 99% で真の値を返す．

実用的な局所差分プライバシのアルゴリズムとして，Google Chrome に実装されている RAPPOR [10]，Apple スマートデバイスに入力予測のために導入されている Private Count Mean Sketch [5] などが挙げられる．文献 [7] では，局所差分プライバシの様々な方式について有効性の比較評価を行っている．本稿では，RAPPOR で検証を行う．RAPPOR はサーバに送信する情報を m 次元ベクトル v で表現し，v の各要素に対してランダム化応答に基づく確率メカニズムを適用する．サーバ側では多数のクライアントから送信されたランダム化ベクトルを集計し，各ビットが 1 である割合を推定する．

文字列などの大きなドメインをもつデータを扱う場合はドメインの各要素をビット列に割り当てる方法では m の値が非現実になるため，RAPPOR では Bloom フィルタを用いて次元数を圧縮する．この場合，サーバ側ではドメインのすべての要素を対象とするのではなく，出現頻度の高い上位 k 個の要素とその頻度を推測する heavy hitters 問題 [6] を扱うことが多い．

局所差分プライバシはサーバ側でノイズが蓄積するため，実用的な有用性を達成するためには一般に差分プライバシ適用時よりも多くのデータ数が必要である．

2.2.3　差分プライバシと機械学習

教師あり機械学習は入力 $\boldsymbol{x} \in \mathcal{X}$ からラベル $y \in \mathcal{Y}$ を予測する関数 $f : \mathcal{X} \to \mathcal{Y}$ として定式化される．学習時は，訓練データ $D = \{(\boldsymbol{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} \mid i = 1, \ldots, n\}$ を用いて，正解ラベルとモデルの予測値の差によって定められた損失関数 $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{x}, y)$ の期待値を最小にするようにモデルパラメータ $\boldsymbol{\theta}$ の最適化を行う．損失関数として，回帰問題の場合は平均二乗誤差，分類問題の場合は交差エントロピーがよく使われる．損失を最小化する代表的な手法として，確率的勾配降下法（stochastic gradient descent; SGD）がある．SGD では，ミニバッチ $B \subset D$ をランダムに選択し，ミニバッチごとに次式を用いて $\boldsymbol{\theta}$ を繰り返し更新する．

$$\boldsymbol{\theta_{i+1}} \leftarrow \boldsymbol{\theta_i} - \eta \cdot \frac{1}{|B|} \sum_{(\boldsymbol{x}, y) \in B} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta_i}; \boldsymbol{x}, y)$$

ここで，η は学習率，$\nabla_{\boldsymbol{\theta}}$ はパラメータの勾配である．

SGD に差分プライバシを適用した DP-SGD [3] では上のパラメータ更新式に正規分布に従うノイズを加えてプライバシ保護を行う．具体的な更新式は以下の通りである．

$$\boldsymbol{\theta_{i+1}} \leftarrow \boldsymbol{\theta_i} - \eta \cdot \frac{1}{|B|} \left(\sum_{(\boldsymbol{x}, y) \in B} \mathrm{clip}(\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta_i}; \boldsymbol{x}, y), C) + \mathcal{N}(0, \sigma^2 C^2 \boldsymbol{I}) \right)$$

ここで，C は勾配の L2 ノルムの閾値であり，

$$\mathrm{clip}(\boldsymbol{v}, C) = \boldsymbol{v} \cdot \min\left(1, \frac{C}{||\boldsymbol{v}||_2}\right)$$

である．DP-SGD が満たす ϵ の上界は文献 [3] に示されている．

機械学習におけるプライバシ侵害として，学習済みモデルから特定の個人のデータが訓練データ D に含まれるかどうかを推測するメンバーシップ推論攻撃がある．ブラックボックス型の攻撃では，攻撃者が対象のモデルに入力を与え，その出力を観察することで入力が訓練データに含まれているか否かを推論する．例えば，成績の悪い学生のみを対象として作成した予測モデルを公開する場合，そのモデルの訓練データとして自身のデータが含まれることを第三者に知られることはプライバシ侵害につながる．モデルが差分プライバシを満たしていれば，メンバーシップ推論攻撃に対する防御として有効である．文献 [15] では，画像データを用いた差分プライバシを満たすモデルに対するメンバーシップ推論攻撃の詳細について報告されている．

3　差分プライバシ適用による精度の比較

以下では，いくつかの分析手法に対して，差分プライバシを適用した場合と適用しなかった場合についてプライバシ保護と有用性への影響を比較検証する．

3.1　ヒストグラム

3.1.1　差分プライバシ

データセットとして Kaggle Students Performance in Exams [*1] を用いた．このデータセットは 1,000 人分の学生の属性と 3 科目の点数からなる．ここでは，この中からランダムに選択した 50 人分の数学の点数のみを用いた．平均点は 62.8，標準偏差は 15.2，最低点は 18，最高点は 97 である．各点数を 80 以上は A，70 以上 80 未満は B，60 以上 70 未満は C，50 以上 60 未満は D，40 以上 50 未満は E，40 未満は F の 6 段階で評価し，段階別のヒストグラムを作成する．

このヒストグラムに対して Laplace メカニズムによる差分プライバシを適用する．後処理として，ノイズを加えた結果ビンの値が負になる場合は 0 に補正する処理を施した．以降，実装はすべて Google Colaboratory の Python 3.7 上で行い，差分プライバシのライブラリには PyDP [*2] を使用した．有用性の評価は元の分布との平均二乗誤差（MSE）で行う．ϵ を 0.05 から 1 まで 20 段階で変化させたときの MSE の値を図 1 に示す．また，例として，$\epsilon = 0.25$ のときの差分プライバシ適用前後のヒストグラム（MSE=5.2）を図 2 に示す．

MSE の許容範囲は目的によるが，16 より大きいときは 1 つのビンでの値の差が平均的に 4 以上となり，$n = 50$ の 6 段階評価の分布としては特に分布の裾の部分において実用上の誤差が大きいと考える[*3]．プライバシ保護について，図 2 右側の差分プライバシ適用後のヒストグラムが元のヒストグラムから得られる確率と，元のヒストグラムから 1 つのデータを除いたものから得られる確率の比は最大でも $e^{0.25} \simeq 1.28$ 倍であることが保証される．

[*1] https://www.kaggle.com/spscientist/students-performance-in-exams

[*2] https://github.com/OpenMined/PyDP

[*3] なお，計数問合せ（count query）の感度は 1 なので，2.2.1 章で述べた出力誤差の理論上の上界は確率 95% で $\frac{1}{0.25} \ln \frac{1}{0.05} \simeq 12.0$ である．

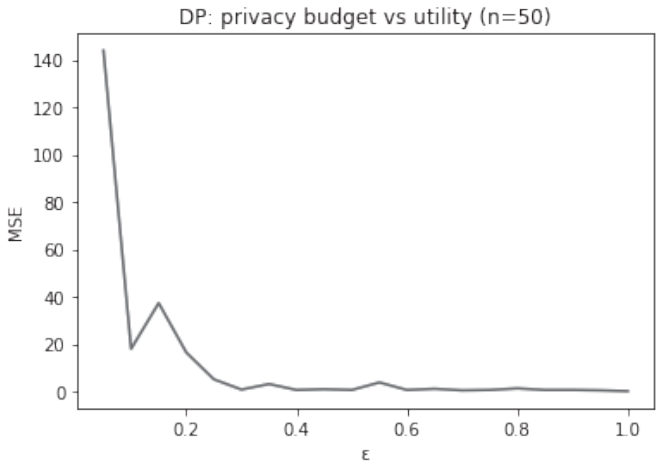

図 1　成績分布に対する差分プライバシの有用性評価

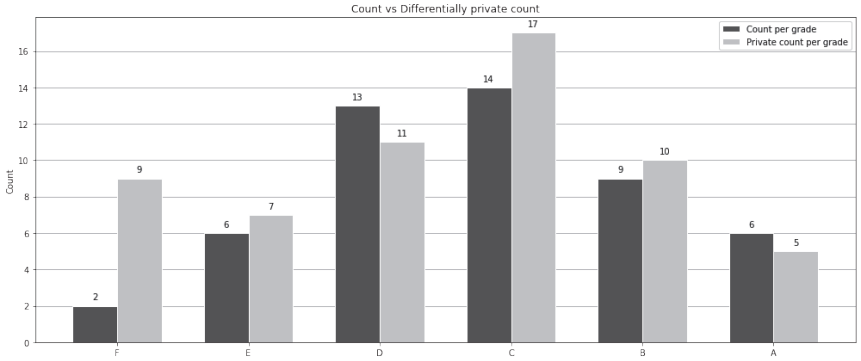

図 2　$\epsilon = 0.25$ の差分プライバシ適用前後のヒストグラム

3.1.2　局所差分プライバシ

データセットとして，Open University Learning Analytics dataset [13] に含まれている，オンライン学習環境上での教材へのアクセス履歴データを用いた．このうち，5 万件のアクセスに対して局所差分プライバシを適用してアクセス頻度を推定する．RAPPOR の実装には pure-ldp [*4] を使用した．Bloom フィルタの次元数は 128 とした．

[*4] `https://github.com/Samuel-Maddock/pure-LDP`

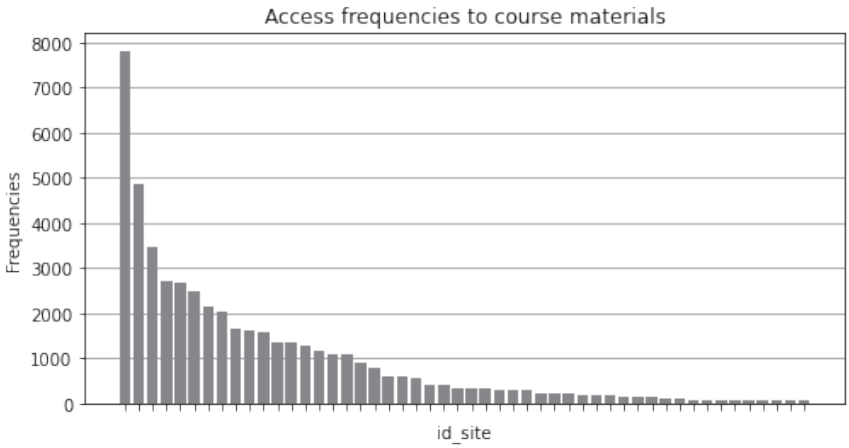

図 3　教材ごとのアクセス頻度

元データにおける教材ごとのアクセス頻度は図 3 のような Zipf の法則に従った分布となっている．横軸は教材 ID であるが，表示の簡略のため省略している．ϵ を 0.5 から 5 まで 10 段階で変化させたときの MSE の値を図 4 に示す．また，例として，$\epsilon = 3$ のときの局所差分プライバシ適用前後のヒストグラムを図 5 に示す．RAPPOR では，$\epsilon = 3$ は確率 0.32 でビット反転が起きることに相当する．$\epsilon = 5$ では MSE は低い値になっているが，確率 0.075 でしかビット反転されないためプライバシはほとんど保護されない．

3.2　ナイーブベイズ分類器

前述の Students Performance in Exams データセットを用いて，学生の 3 科目の点数から性別を予測する問題を考える．ここでは，あるデータがどのカテゴリに属するかを確率的に求める機械学習の 1 つであるナイーブベイズ分類器を用いる．ナイーブベイズ分類器の学習では，クラスを $C = \{c_1, \ldots, c_k\}$，入力変数を $\boldsymbol{x} = \langle x_1, \ldots, x_n \rangle$ としたとき，次の尤度関数

$$L(\boldsymbol{x}, C) = \prod_{j=1}^{k} \left[P(C_j) \prod_{i=1}^{n} P(x_i | C_j) \right]$$

を最大化することによって確率分布のパラメータ，Gaussian 分布を仮定する場合は μ_j と σ_j を導出する．ナイーブベイズ分類器に差分プライバシを適用した手法 [18] では，μ_j と σ_j に対する感度を求め，推定されたパラメータに感度を基にした Laplace ノイズを加

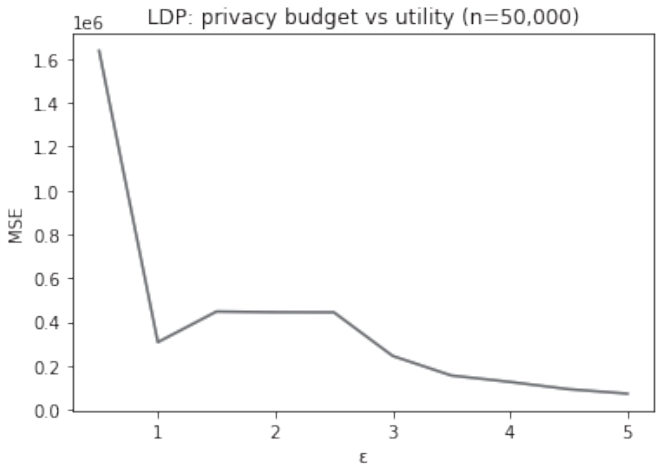

図 4　教材アクセス回数に対する局所差分プライバシの有用性評価

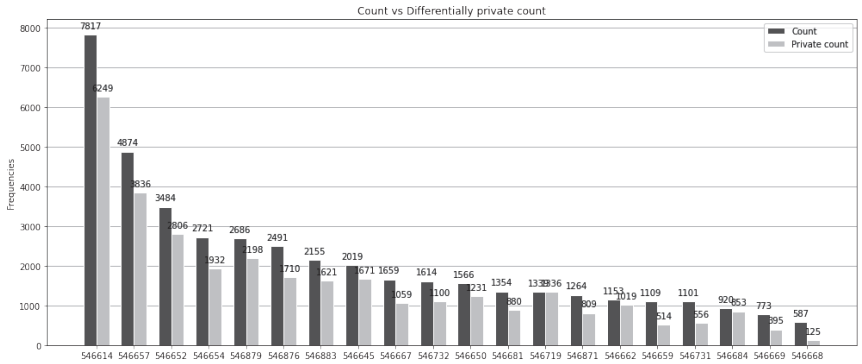

図 5　$\epsilon = 3$ の局所差分プライバシ適用前後のヒストグラム

えることによって ϵ-差分プライバシを達成する．

本データセット中の女性は 518 名，男性 482 名である．このうち，700 件を訓練データ，300 件を検証データとして分類器を構築および評価する．評価指標には正解率を使用し，交差検証を行って算出した．$\epsilon = [10.0, 1.0, 0.1, 0.05, 0.01, 0.005, 0.001]$ としたときの正解率の推移を図 6 に示す．scikit-learn を用いた差分プライバシを適用しないナイーブベイズ分類器による正解率 0.69 をベースラインとして示している．$\epsilon = 1.0$ での検証時の混同行列の例を表 2 に示す．結果として，$n = 1,000$ では感度が十分低い値にならず

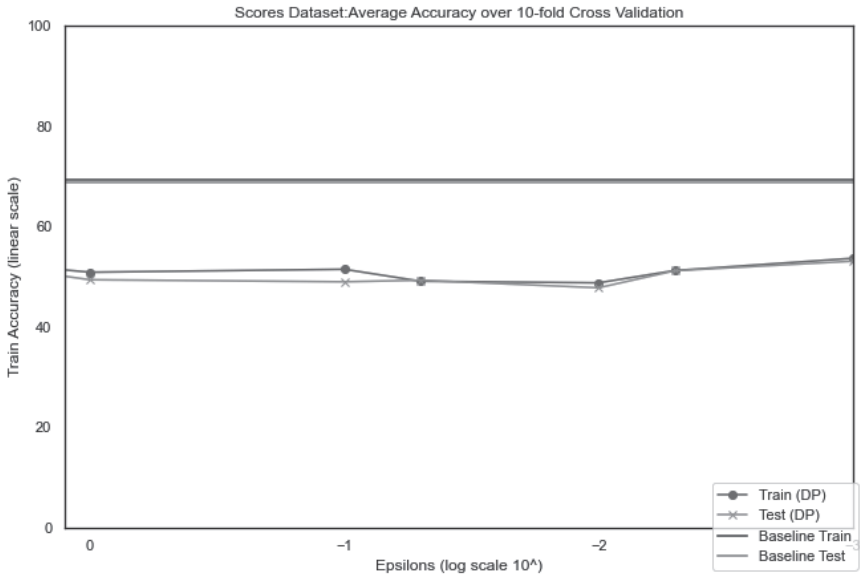

図 6　差分プライバシを適用したナイーブベイズ分類器の有用性の評価

表 2　ナイーブベイズ分類器による性別予測の混同行列の例

ベースライン

実際の値＼予測値	女性	男性
女性	37	13
男性	19	31

$\epsilon = 1.0$

実際の値＼予測値	女性	男性
女性	41	9
男性	32	18

ノイズの影響が大きいために $\epsilon = 1.0$ でもランダムとほぼ変わらない正解率となっており，実用的な精度を得るためにはより大規模なデータが必要であると言える．

3.3　ランダムフォレスト

データサイズの影響を確認するために，より大規模なデータセット [8] を用いて前節と同じく複数科目の点数から性別を予測する問題を考える．本データセットから女性，男性のデータをともに 5,000 ずつ抽出した．このうち，9,000 件を訓練データ，1,000 件を検証データとして決定木を構築および評価する．構築された決定木の例を図 7 に示す．差

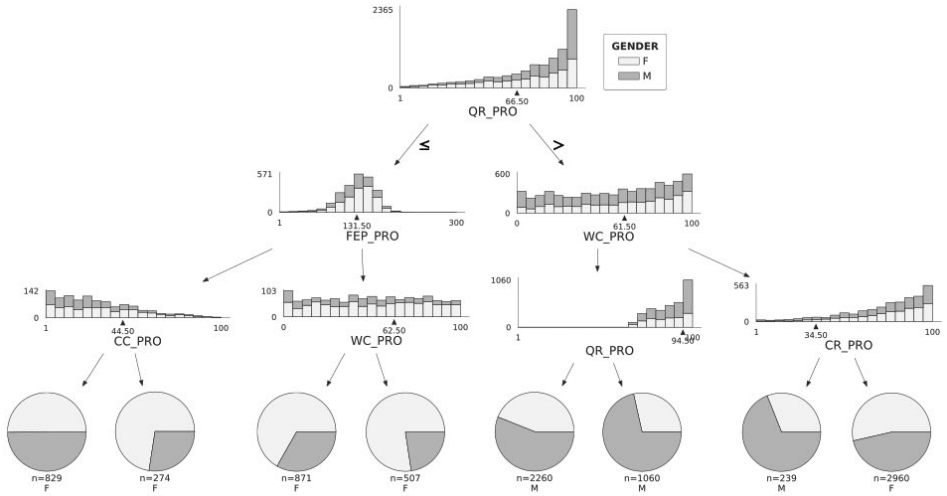

図 7　ランダムフォレストによる性別予測の決定木の例

分プライバシを適用しないランダムフォレストによる正解率は 0.63 である．

ここでは，モデル作成者とデータ所有者が異なり，モデル作成者がデータ所有者に対して差分プライバシを適用した問合せ結果を要求してモデルを構築する，という問題設定を考える．実装には Smooth Random Trees [*5] を使用した．本手法は決定木の葉ノードに対して最頻のラベルと次に頻度の高いラベルとの差を考慮することによって感度を低減化した問合せを行い，結果としてクラスラベルのみを得ることによって消費するプライバシーコストを削減している [11]．差分プライバシは葉ノードのクラスラベルに指数メカニズムを使用したノイズを加えることによって実現される．$\epsilon = [10.0, 1.0, 0.1, 0.05, 0.01, 0.005, 0.001]$ としたときの正解率の推移を図 8 に示す．$\epsilon = 1.0$ での検証時の混同行列の例を表 3 に示す．$\epsilon = 1.0$ 未満ではランダムとほぼ変わらない結果となったが，$\epsilon = 1.0$ のときの正解率は 0.57 であり，ベースラインからの低下の幅は許容範囲にあると考えられる．

3.4　ニューラルネットワーク

データセットとして Kaggle Students' Academic Performance Dataset [4] を用いる．このデータセットは 460 人の学生について，デモグラフィック，学歴，行動特性などから

[*5] https://github.com/sam-fletcher/Smooth_Random_Trees

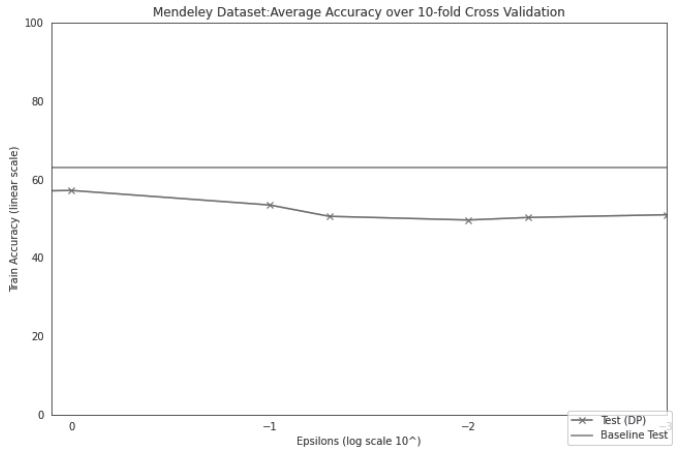

図 8　差分プライバシを適用したランダムフォレストの有用性の評価

表 3　ランダムフォレストによる性別予測の混同行列の例

ベースライン

実際の値＼予測値	女性	男性
女性	303	189
男性	180	328

$\epsilon = 1.0$

実際の値＼予測値	女性	男性
女性	329	305
男性	273	404

なる 16 の属性と成績を 3 段階で示す 1 つのクラス属性からなる．ここでは，16 の説明変数から成績を予測する問題を深層学習を使って解くことを考える．

ベースラインとして構築した分類器は 3 層の全結合型ニューラルネットワークで，ニューロン数は 32，64，128，活性化関数として中間層では ReLU，出力層ではソフトマックスを使用した．これに対し，差分プライバシを適用した同様の分類器を構築し，学習率を 0.15 として正解率の比較を行った．実装には TensorFlow，DP-SGD のライブラリは dpsgd-optimizer [*6] を使用した．

エポック数 50 で学習させたときの正解率と消費したプライバシ予算の推移を図 9 に示す．正解率はエポック数 20 で SGD が 0.91，DP-SGD が 0.57，その時点で消費したプライバシ予算は 8.8 となった．本実験でも $n = 500$ 程度では十分な有用性が得られたと

[*6] https://github.com/thecml/dpsgd-optimizer

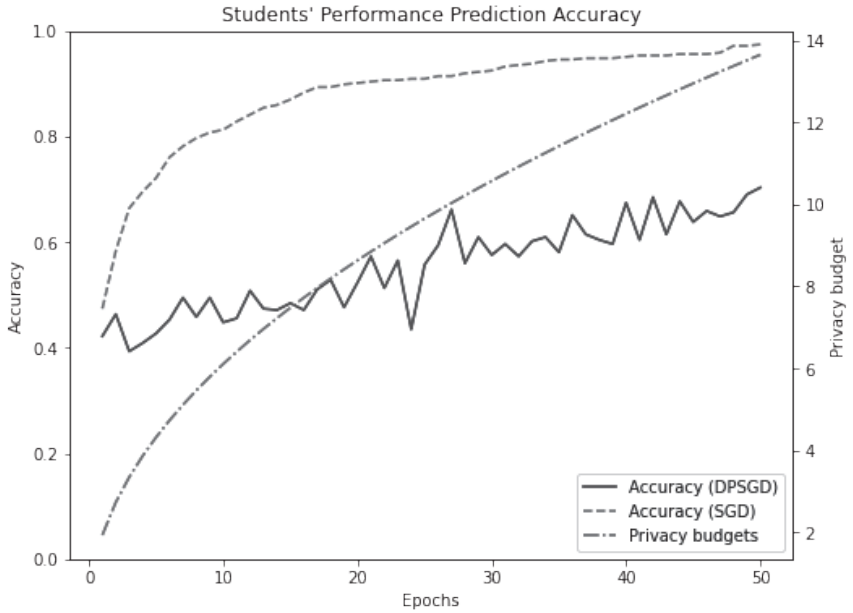

図 9　差分プライバシを適用したニューラルネットワークによる成績予測

は言えないが，プライバシ予算に応じて精度が上がる傾向にあり，n が十分大きければより実用的な有用性を達成できるものと考えられる．

4　まとめ

本稿では，公開されている教育データセットを用いて，いくつかのデータ分析手法に対して，学習者のプライバシを保護する差分プライバシの適用可能性について検討した．

本稿で採用した差分プライバシ技術はベースラインの手法であり，プライバシ予算や有用性指標に関しては様々な改良が行われている．今後は，より洗練化された手法や局所差分プライバシを利用した連合学習，合成データ生成技術に基づく有用性の検証を行う予定である．

参考文献

[1] 緒方広明, 藤村直美: 大学教育におけるラーニングアナリティクスのための情報基盤システムの構築, 情報処理学会論文誌 教育とコンピュータ (TCE), Vol.3, No.2, pp.1-7 (2017).

[2] 佐久間淳: データ解析におけるプライバシー保護, 講談社, 2016.

[3] M. Abadi, A. Chu, I.J. Goodfellow, H. Brendan McMahan, I. Mironov, K. Talwar, L. Zhang: Deep Learning with Differential Privacy. CCS 2016: 308-318.

[4] E.A. Amrieh, T. Hamtini, I. Aljarah: Mining educational data to predict student's academic performance using ensemble methods. International Journal of Database Theory and Application, 9(8), 119 - 136 (2016).

[5] Apple: Learning with Privacy at Scale, Dec. 2017. `https://machinelearning.apple.com/research/learning-with-privacy-at-scale`

[6] R. Bassily, K. Nissim, U. Stemmer, A. Guha Thakurta: Practical Locally Private Heavy Hitters. NIPS 2017: 2288-2296.

[7] G. Cormode, S. Maddock, C. Maple: Frequency Estimation under Local Differential Privacy. Proc. VLDB Endow. 14(11): 2046-2058 (2021).

[8] E. Delahoz-Dominguez, R. Zuluaga, T. Fontalvo-Herrera: Dataset of academic performance evolution for engineering students, Data in Brief, 30, 105537, 2020.

[9] C. Dwork: Differential Privacy. ICALP (2) 2006: 1-12.

[10] U. Erlingsson, V. Pihur, A. Korolova: RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. CCS 2014: 1054-1067.

[11] S. Fletcher, M.Z. Islam: Differentially Private Random Decision Forests using Smooth Sensitivity. CoRR abs/1606.03572 (2016).

[12] A. Hernandez-Blanco, B. Herrera-Flores, D. Tomas, B. Navarro-Colorado: A Systematic Review of Deep Learning Approaches to Educational Data Mining. Complex. 2019: 1306039:1-1306039:22 (2019).

[13] J. Kuzilek , M. Hlosta, Z. Zdrahal: Open University Learning Analytics dataset Sci. Data 4:170171 doi: 10.1038/sdata.2017.171 (2017).

[14] M.C. Mihaescu, P.S. Popescu: Review on publicly available datasets for educational data mining. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 11(3) (2021).

[15] M.A. Rahman, T. Rahman, R. Laganiere, N. Mohammed: Membership Inference Attack against Differentially Private Deep Learning Model. Trans. Data Priv. 11(1): 61-79 (2018).

[16] C. Romero, S. Ventura: Educational data mining and learning analytics: An updated survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 10(3) (2020).

[17] S.A. Salloum, M. Alshurideh, A. Elnagar, K. Shaalan: Mining in Educational Data: Review and Future Directions. AICV 2020: 92-102.

[18] J. Vaidya, B. Shafiq, A. Basu, Y. Hong: Differentially Private Naive Bayes Classification. Web Intelligence 2013: 571-576

(本学准教授)