［差分プライバシ］

1 データプライバシ保護技術と差分プライバシ

古川　諒 | NEC

パーソナルデータ活用とデータプライバシ保護技術

近年 ICT の進化，特にスマートデバイス・IoT（Internet of Things）機器の普及に伴い，個人にかかわる情報（パーソナルデータ）は，氏名や年齢，住所，ゲノムといった個人に固有の情報のみならず，購買履歴や動画や Web サイトの閲覧履歴，位置情報などの個人の行動を表す履歴情報や，顔画像・動画などのリッチメディアなど多岐に渡り，さまざまな事業者や組織によって収集されるようになってきた．深層学習の進化に端を発するビッグデータ活用の機運の高まりはこれらのパーソナルデータの活用に対してもおよび，社会的・産業的に有用な知見を得ることが期待されている．

典型的なパーソナルデータ活用といえば企業・組織が自分たちで収集したデータのみを用いて，サービス向上や顧客ごとにカスタマイズされたサービス（パーソナライズドサービス）を提供するものである．このような活用方法はパーソナルデータおよびその分析結果が収集した主体の中にとどまるため，有効な同意に基づいてパーソナルデータを収集できていればプライバシ侵害につながることは少ない．

一方で，国政調査のようにパーソナルデータから算出した統計値を公開したり，近年では MLaaS (Machine Learning as a Service) のように機械学習モデルを第三者が使えるように公開したり，第三者に分析してもらうためにデータを提供したりといった活用方法もある．このような活用方法では悪意ある第三者によって，提供された情報から元のパーソナルデータの一部が漏洩し，個人に不利益をもたらす危険性がある．たとえば，2009 年に Netflix がレコメンデーションアルゴリズム開発コンテストのためにユーザの映画評価情報を公開したところ，ほかのレビューサイトで公開されていた情報との突合により個人が特定され，個人が隠していた映画視聴履歴が漏洩してしまい，訴訟に発展するといった事件が発生している．

データプライバシ保護技術はこのようなパーソナルデータの活用と保護を両立するための技術として研究されてきた．これらの技術ではパーソナルデータそのものを加工したり，分析結果を加工したりすることで，機微な情報と個人との紐づけを防止しながら活用することを目的としている．

本稿ではこれらのデータプライバシ保護技術についての概観を解説し，本小特集で焦点を当てる差分プライバシの位置づけについて述べる．

データプライバシ保護技術の概観

データプライバシ保護技術の分類の 1 つに「入力プライバシ保護技術」「出力プライバシ保護技術」「秘密計算技術」の 3 つの技術カテゴリに分ける方法がある．本稿ではこの分類をベースに解説を行う．

それぞれの技術カテゴリは以下のように説明できる．

- 入力プライバシ保護技術：分析処理に入力するパーソナルデータを加工することで分析者によるプライバシ侵害のリスクを低減する技術

- 出力プライバシ保護技術：パーソナルデータに対する分析結果からのプライバシ侵害のリスクを低減する技術
- 秘密計算技術：暗号やそれに類する技術により，データ分析者に対してパーソナルデータを秘匿しながら分析を行う技術

図 -1 はデータ保有者が保有するパーソナルデータをデータ収集者が収集し，統計処理や機械学習などの分析処理の実行者であるデータ分析者が収集されたデータを分析し，分析結果を最終的に利用する過程において，各技術カテゴリの関係を示したものである．ここで，データ収集者とデータ分析者が同一であったり，違う人物であったりする場合☆1 もあるし，同様に，データ分析者と分析結果利用者が同一であったり，違う人物であったりする場合☆2 があることに注意されたい．

図中で同じ色の矢印は一連のデータフローを表しており，図中には 4 つの考えられるデータフローを示している．これらのデータフローのうち，フロー①，③はデータ収集者，データ分析者，分析結果利用者がすべて異なる人物でもプライバシ保護が可能なデータフローである．一方でフロー②はデータ収集者とデータ分析者の間でプライバシ保護がされないため，両者が同一の人物であることが求められ，フロー④はデータ分析者に対してパーソナルデータは秘匿されるが分析結果は加工されていないため，分析結果利用者がデータ収集者であることが求められる．

以降の節でそれぞれの技術カテゴリにおいて利用される技法について簡単に述べる．

入力プライバシ保護技術

図 -1 に示したように入力プライバシ保護技術はデータ保有者（個人）またはデータ収集者によって実行される．基本的には入力された個票データを加工し，第三者に公開できる個票データを出力する技術である（図 -2）．

入力プライバシ保護技術に使用される加工技法の代表例を表 -1 に示す．入力プライバシ保護技術はこれらの加工技法を組み合わせて個票データを加工し，公開した個票データから個人と機微な情報を紐づけられてしまうリスクを低減する技術と言える．

出力プライバシ保護技術

出力プライバシ保護技術はデータ分析者が分析処

☆1 この場合，データ収集者とデータ分析者の間にはたとえばデータを売買するような関係がある．

☆2 この場合，データ分析者と分析結果利用者の間にはたとえば分析結果を売買するような関係がある．

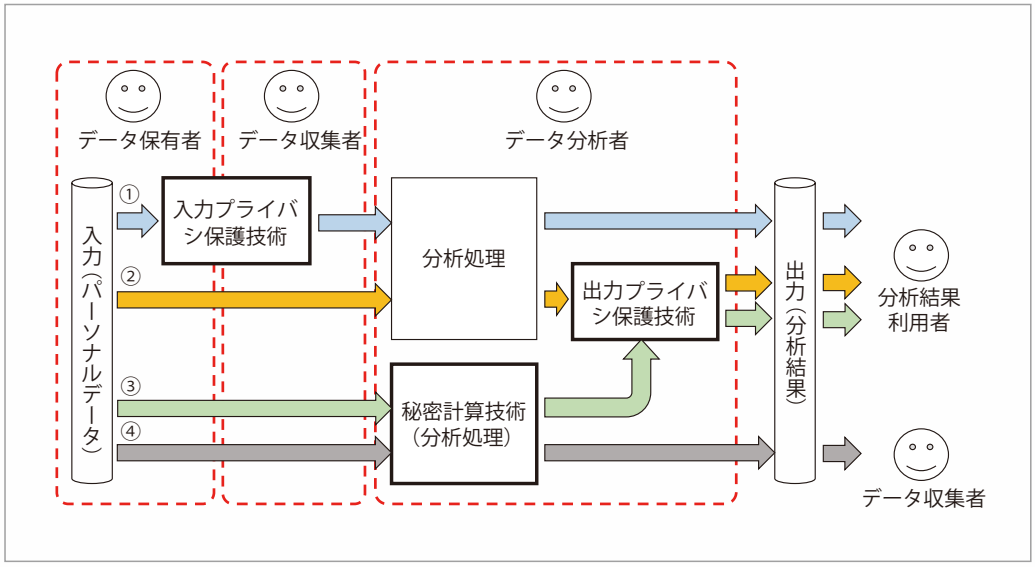

図 -1　プライバシ保護技術カテゴリの関係性

理を行い，分析結果利用者に対して出力する際に実行され，個票データを分析した結果を何らかの形で制御することでプライバシを保護する．

出力プライバシ保護技術では図 -3 のように主に分析結果に乱数的なノイズを付加する技法が用いられる．また，分析結果利用者が知りたい統計情報を指定するようなモデルの場合，複数の統計出力によりプライバシが漏洩しないかを検査し，出力を抑制するクエリ監査と呼ばれる技法もある．

秘密計算技術

秘密計算技術では暗号技術などによりデータを秘匿したまま分析処理を実行する．秘密計算技術は大きく分類して秘密分散を用いた方法，準同型暗号を用いた方法，ハードウェアを用いた方法があり，それぞれ速度や安全性，計算に参加するパーティ数などが異なる．

秘密計算技術は他の 2 つとは異なり，出力される処理結果は元データに対する処理結果と一致する．このため分析結果からのプライバシ侵害を防ぐためには別途出力プライバシ保護技術などと組み合わせる必要がある．

ID	年齢	性別	郵便番号
1	25	男	111-1111
2	30	女	222-2222
3	32	男	333-3333
4	61	女	444-4444

年齢	性別	郵便番号
25-30	男	111-11**
25-30	女	222-2***
32	*	333-****
60以上	*	44*-****

図 -2　入力プライバシ保護技術による加工の例

表 -1　入力プライバシ保護技術の加工技法 [1)]

加工技法	加工内容
削除	列，行，セルなどの単位で属性値を削除する
仮名化	識別子を削除または置き換える
一般化	値を上位の概念や範囲に置き換える
曖昧化	特に大きい，または小さい数値をまとめる（〇〇以上など）
ミクロアグリゲーション	複数のレコードの値をその集合の代表値に置き換える
ノイズ付加	値に乱数的なノイズを加える
データ交換	レコード間で値の（確率的な）入れ替えを行う
疑似データ生成	元データと統計的性質が類似するように人工データを生成する

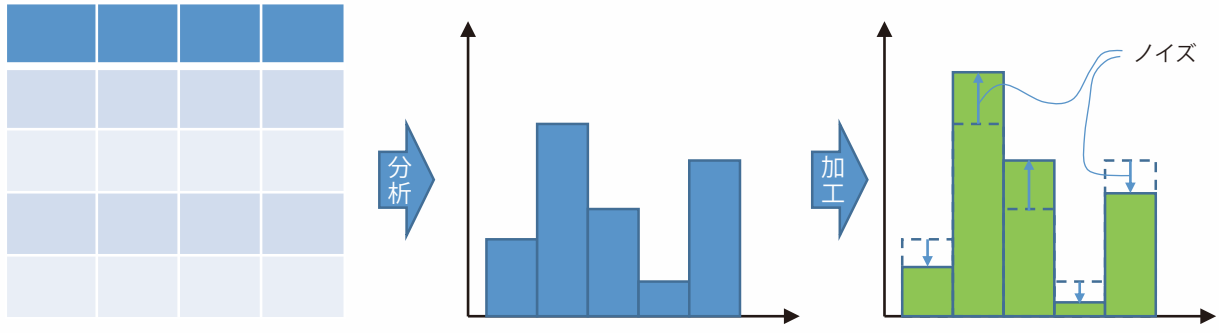

図 -3　出力プライバシ保護技術の加工の例

差分プライバシの位置づけ

差分プライバシの概念

前述のとおりノイズ付加やデータ交換などの乱数を用いた技法（攪乱的技法と呼ぶ）は入力プライバシ保護技術においても出力プライバシ保護技術においても用いられている．しかしながら，「どのようなノイズであればどの程度安全なのか」明確な基準がないことは長らく問題であった．

Dwork によって提案された“差分プライバシ”[2]は，攪乱的技法に対する安全性を数学的に定義した指標である．詳細な定義は後の記事に譲るが，端的に言えば「（ノイズが付加された）出力が 2 つの近しいパーソナルデータ集合の間で区別ができないのであれば安全である」ということを定義した指標である．つまり，差分プライバシを満たせば，2 つの近しいデータ集合の（任意の）差分が出力に影響を与えないようにノイズが加わることになるため，出力から確定的に差分の内容を特定できなくなり，プライバシを保護できるのである（図 -4）．このような定義により，特に攻撃者の前提知識や攻撃方法を仮定せずに，任意の分析方法☆3 に対して安全性を議論できることが注目を浴びている．

差分プライバシは入力プライバシ保護，出力プライバシ保護どちらの指標としても利用されている．つまり，適用対象の処理として統計値や機械学習などにノイズを付加する場合には出力プライバシ保護に，ノイズを付加した疑似データを出力する場合には入力プライバシ保護に使用されることになる．また，近年ではデータ保有者（個人）からパーソナルデータを収集する際に保有者側でノイズを付加するプライバシ保護が注目を集めており，このようなケースでも差分プライバシを利用できる．このケースでは差分プライバシの派生である“局所差分プライバシ”が提案されており，多くの研究がされている．

☆3 実際には差分プライバシを満たすように分析方法を改変する必要があり，改変が困難な場合もある．

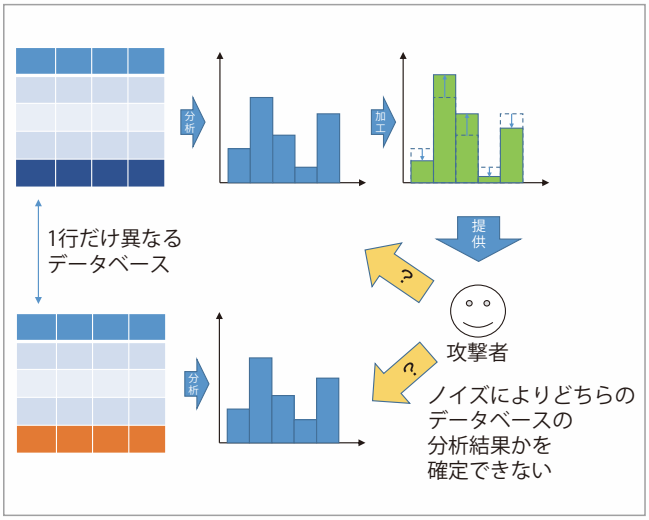

図 -4　差分プライバシのイメージ

差分プライバシと他指標の比較

ここで，入力プライバシ保護技術においてよく用いられる k- 匿名性[3]と比較をしてみよう．k- 匿名性とは，加工された個票データにおいて個人の識別に使用できる属性（たとえば年齢，性別，郵便番号などであり，これらを総称して準識別子と呼ばれる）の組合せが少なくとも k 人以上同じである場合 k- 匿名性を満たすという指標である．

表 -2 はそれぞれの指標に対して対応する技術カテゴや指標が対象としている加工技法，指標を利用できるプライバシ保護の実行者や攻撃者の仮定をまとめている．差分プライバシに対しては適用対象とする処理の観点で，前節で述べた「統計分析・機械学習等」「疑似データ生成等」「個人からのデータ収集」と 3 つに分けてその特性を詳細に示している．「個人からのデータ収集」のケースは前述した局所差分プライバシが適用されるケースにあたる．

表 -2 から分かるように，k- 匿名性がデータ収集者の実行する非攪乱的な加工に対してのみ適用できるのとは異なり，差分プライバシはデータ分析者，データ収集者，データ保有者が実行する攪乱的な加工に対して適用できることが分かる．また，k- 匿名性は攻撃者が前述した準識別子を知っており，出

力された個票データに含まれる準識別子との一致を見ることで個人を特定するような攻撃に対する安全性を定義しているが，差分プライバシは攻撃者がどのような知識を持っているか，どのような攻撃を実行するかによらず一元的に安全性を議論できる．

このように，差分プライバシはデータプライバシ保護技術の中で，広範な攪乱的技法に適用可能な安全性指標であると位置づけられる．以上の整理が以降の記事の理解を助けるものであれば幸いである．

参考文献

1) 内閣官房：技術検討ワーキンググループ報告書（第5回），https://www.kantei.go.jp/ jp/singi/it2/pd/dai5/siryou2-1.pdf
2) Dwork, C. : Differential Privacy, Proc. 33rd Intl. Conf. Automata, Languages and Programming - Volume PartII, LNCS, Vol.4052, pp.1-12 (2006).
3) Sweeney, L. : k-anonymity : A Model for Protecting Privacy, International Journal on Uncertainty, Fuzzi-ness and Knowledge-based Systems,10(5), pp. 557-570 (2002).

（2020年2月26日受付）

■古川　諒（正会員）　rfurukawa@nec.com

2008年東京工業大学総合理工学研究科博士前期課程修了．同年NEC入社．以来，アクセス制御，プライバシ保護，ブロックチェーンの研究に従事．

表-2　*k*-匿名性と差分プライバシの比較

指標	*k*-匿名性	差分プライバシ		
		統計分析・機械学習等	疑似データ生成等	個人からのデータ収集（局所差分プライバシ）
技術カテゴリ	入力プライバシ	出力プライバシ	入力プライバシ	入力プライバシ
対象とする加工技法	一般化・削除・曖昧化などの非攪乱的技法	ノイズ付加，データ交換などの攪乱的技法		
保護の実行者	データ収集者	データ分析者	データ収集者	データ保有者
攻撃者の知識	準識別子	仮定しない		
対処できる攻撃	個人の特定	仮定しない		