特　集　「パーソナルデータの保護と活用」

パーソナルデータ秘匿化技術のサービス活用

Utilization of Personal Data Protection Technologies for Services.

古川　諒　NEC セキュリティ研究所

Ryo Furukawa　NEC Security Research Labs.
rfurukawa@nec.com

佐古　和恵　早稲田大学　基幹理工学部

Kazue Sako　Waseda Univ. School of Fundamental Science and Engineering
KazueSako@aoni.waseda.jp

Keywords: パーソナルデータ活用, プライバシー保護, k-匿名化, マルチパーティ計算

1. はじめに

近年，ICT 技術の進化・普及に伴い，氏名や生年月日，住所や個人の行動履歴といった個人に関わる情報が様々なサービスによって収集されている．これらの個人に関わる情報は総称してパーソナルデータと呼ばれる．収集されたパーソナルデータはそれぞれのサービスで縦横に活用されている．例えば，EC サイトにおける購買履歴に基づいた商品を推薦するレコメンデーションサービスは典型的な例である．

このようなパーソナルデータを活用したサービスによって個人に対して新しい価値を提供できる．また，都市開発や医療分野といった領域でもパーソナルデータの価値は重要になりつつあり，公益性の高い活用も見込まれている．

一方でパーソナルデータは個人のプライバシーを含んでおり，個人からの忌避感や法規制により活用に壁があるのも事実である．EU では実際に GDPR（EU 2016）違反を犯し多額の罰金を支払ったケースも発生している（CNIL 2019，ICO 2019）．

これらの壁は特にデータを収集した主体とデータを活用したい主体が異なる場合に顕著になる．このような場合でもパーソナルデータを取得するための目的をしっかりと個人に説明し，同意を取得することは重要である．それに加え，個人が不利益を受ける危険性を低減しながらパーソナルデータの活用を可能とする技術（ここではパーソナルデータ秘匿化技術と称する）を用いることでひとつのサービスに閉じない活用が可能になる．

本稿では，パーソナルデータ秘匿化技術のうち 2 つの技術について紹介を行う．紹介する技術は「データ匿名化技術」と「マルチパーティ計算技術」である．これらの技術が社会的に受容されて汎用的に使われるためには今後も議論が必要だが，読者がパーソナルデータの活用を考えたときの選択肢のひとつとなれば幸いである．

本稿は以下のように構成される．2 章ではパーソナルデータの活用を形態分けし，紹介する技術について簡単な位置づけを述べる．3 章ではデータ匿名化技術について詳細に紹介する，4 章ではマルチパーティ計算技術について紹介する．5 章はまとめである．

2. パーソナルデータ秘匿化技術の適用ケース例

パーソナルデータ秘匿化技術をサービスに適用する際，誰がデータを収集し，誰が分析し，その結果を誰が活用するかといった，利用の形態によって適する技術が異なる．そこで本章では本稿で紹介する 2 つの技術が適する活用の形態について述べる．

2.1　パーソナルデータの第三者提供による活用

パーソナルデータの第三者提供による活用とは，パーソナルデータを収集した事業者（データ収集者）で

はない事業者（データ活用者）によってデータが分析され，活用されるケースである．例えば．位置情報サービス事業者が収集した人の位置情報をマーケティング会社へ提供し，商圏分析に役立てたり，EC サイトの購買情報をメーカーへ提供し，製品分析や次期製品開発に役立てるといった事例が該当する．

この形での活用を安全に行うために適した技術は「**データ匿名化技術**」である．

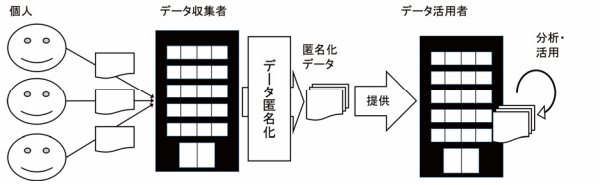

図 1 パーソナルデータの第三者提供

この活用ケースでは，図 1 に示したように，データ収集者が自分たちで収集したパーソナルデータに対してデータ匿名化技術を適用して匿名化データを作成し，その匿名化データをデータ活用者へ提供する．そして，データ活用者は匿名化データに対して統計分析などを行い，自分の事業に活用する．このようにパーソナルデータを安全に活用することで新たな社会的な価値を生み出すのである．

近年では個人情報保護法の改正により，パーソナルデータを匿名加工情報と呼ばれる個人を特定できる情報を排除したデータに加工することで個人の同意なく第三者提供が可能である．また，次世代医療基盤法の施行によって認定事業者により加工された匿名加工医療情報の第三者提供が認められるなど，第三者提供によるパーソナルデータ活用の可能性は高まっている．

また，このケースでは提供したデータは提供先で自由に使うことができるため，提供先で試行錯誤しながら分析できるといった利点がある．その分自由に悪用できるということでもあるため，データに対しては十分な加工が必要となり，分析の正確さが損なわれることには注意されたい．

2.2 複数事業者が持つデータの集約による活用

複数事業者が持つデータの集約による活用とは，パーソナルデータを保有する複数の事業者の間でお互いが保有するデータを合わせた分析をすることで，データから新たな知見を得るケースである．

このような活用を安全に行うために適した技術は「**マルチパーティ計算技術**」である．

この活用ケースでは複数のデータ保有者兼データ活用者が，お互いが持つデータを秘匿した形でマルチパーティ計算を実施し，出力として分析結果を受け取る．そして，この分析結果をそれぞれで自分の事業に活用したり，公表したりすることで社会的な価値を生み出すのである．この活用ケースの事例として，ボストン地域の企業の賃金情報を秘匿したまま集計し，男女の間で 23%の賃金の差があることを明らかにした例がある（Boston Univ 2019）．また，各企業のデータを明かすことなく業界全体の売り上げなどに関わる統計分析を行うといった事例もあるだろう．

なお，マルチパーティ計算はデータ保有者自身が参加者となり計算を実行することで研究が進められてきたが，参加者が多いほど計算が複雑になるため，あらかじめ定められた参加者に委託してマルチパーティ計算を実施する形式もある（図 2）．例えば，上記のボストンの事例では，ボストン大学のマルチパーティ計算プラットフォームが利用された．本稿の 4 章で紹介するのはこちらのケースとなる．

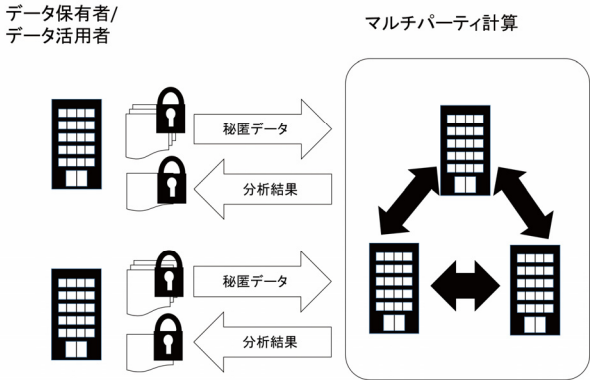

図 2 複数事業者が持つデータの活用

マルチパーティ計算技術はデータ匿名化技術と異なり，入力されるデータの精度が変わらないので，分析の正確さを損なうことなくパーソナルデータを活用できる．その反面，分析処理ごとにマルチパーティ計算アルゴリズムの実装や，その分析を行うための調整が各パーティ間で必要になる．加えて，分析に用いる処理によっては，個人情報が出力されてしまうことから，どのような分析を実施するのかを明示した上で個人から合意を取得しておく必要があり，それ以外の分析は自由に行えない．

3. データ匿名化技術

データ匿名化技術とは，データを加工することにより，個人を特定する手がかりがないようにする技術の総称である．本章ではデータ匿名化技術の中でも特に有名なk-匿名化技術（Sweeney 2002）について紹介する．

k-匿名化技術とは，提供するパーソナルデータの集合が，集合として k-匿名性と呼ばれる指標を満たすようにパーソナルデータを加工する技術である．

3.1 k-匿名性

3.1.1 取り扱うパーソナルデータの形式

k-匿名性は，複数の個人に関する属性情報を記録したデータベースを対象として定義される．このデータベースのことを個票データといい，図 3 に例示している．図 3 では，1 行（1 レコードという）で，一人の患者の属性情報，すなわち，「患者番号」「名前」「ZIP コード」「年齢」「国籍」と「病状」が書かれている．なお，ここで「患者番号」「名前」「ZIP コード」「年齢」「国籍」と「病状」は属性であり，具体的な「000001」「Alice」「13068」「28」「ロシア」「心臓病」などはその属性値であると表現する．

患者番号	名前	ZIPコード	年齢	国籍	病状
000001	Alice	13068	28	ロシア	心臓病
000001	Alice	13068	28	ロシア	皮膚病
000002	Bob	13068	29	アメリカ	感染病
000002	Bob	13068	29	アメリカ	風邪
000003	Carol	13053	21	日本	心臓病
000004	Dave	13053	23	アメリカ	感染病
000005	Ellen	14853	31	アメリカ	風邪
000006	Frank	14853	37	インド	がん
000007	George	14850	36	日本	風邪
000008	Harris	14850	35	アメリカ	がん

識別子　　準識別子　　センシティブ情報

図 3 個票データの例

さらに図 3 に基づいて説明をする．この個票データを疫学的な統計において研究者に提供すると個人の「名前」や「病状」などの情報が記載されているので，問題となる．そこで，データ匿名化技術を用いてこの個票データから個人を類推できないように属性値を加工する．そこで，この個票データにおいて，どの属性の属性値をどういう方針で加工するかを，受け取った先でどのようにデータを活用するかを考慮しながら決めていく必要がある．

ある人がどんな病状を持つかというのはその個人にとっては知られたくない「センシティブ属性」にあたる．一方で，加工された個票データを疫学の研究者が活用するシーンを想定すると，病状に関する属性値は重要な意味を持つ．そこで，「誰が」その属性を持つかという個人を特定できそうな情報を加工することを考える．例えば，「名前」のように，ひとつのサービスに閉じずに使用される社会的な個人の識別情報はまず個票データから削除する必要がある．このような識別子の例としては他にはメールアドレス，マイナンバー番号などが該当する．

「患者番号」のようにひとつのサービスに閉じて利用される情報は，社会的な識別子とは異なり，その番号が外に漏洩してもただちに本人が特定されることはない．一方で，同じ患者番号の人が別のレコードで追加の病状が記載されている場合には，分析を実施するために連結して有効な情報になる．そこで，このような情報は，一意性を保ちつつ，別の番号に変換されることがある．

このように個人識別情報に対して処理をしたとしても，図 3 のように，個票データにおいて「ZIP コード」や「年齢」から本人が一意に特定されてしまうことがある．例えば国籍がアメリカで 29 歳の患者は Bob だけである，ということが分かってしまうかもしれない．

k-匿名性を考えるにあたって，このような複数の属性の属性値を組み合わせることで個人を一意に識別できる可能性が高い属性を「準識別子」と呼ぶ．k-匿名化の技術はこの準識別子に属する属性値を加工することによって本人を特定しにくい個票データを生成することを目的とする．

3.1.2 k-匿名性の定義

k-匿名性は準識別子による個人の識別の困難さを表す指標である．k-匿名性は以下のように定義できる．

定義 1. k-匿名性 : 個票データの中で，準識別子として扱う属性の集合について同一の属性値の組み合わせを持つレコードが少なくとも k 人以上必ず存在するとき，その個票データは k-匿名性を満たすという．

ここで，k はセキュリティパラメータであり，k が大きければ大きいほど個票データから個人の識別が困難になる．k-匿名性を図示すると図 4 のようになる．k-匿名性を満たした個票データからは，ある個人の準識別子（図では年齢，国籍，Zip コードを知っていたとしても，その個人のレコードを特定することができず，センシティブな属性（病状）を推論しづらくなる．これによりプライバシーの保護につながるのである．

No.	ZIPコード	年齢	国籍	病状	
1	13068	28-29	*	心臓病	k=2
2	13068	28-29	*	感染病	
3	13053	21-23	*	心臓病	k=2
4	13053	21-23	*	感染病	
5	14853	31-37	*	風邪	k=2
6	14853	31-37	*	がん	
7	14850	35-36	*	風邪	k=2
8	14850	35-36	*	がん	

図 4 k-匿名性の例

3.2 k-匿名化技術

サービス事業者が収集した状態の個票データが k-匿名性を満たしていることはほぼない．このため，個票データの各属性値を加工して k-匿名性を満たす必要がある．加工された個票データを用いると分析結果に誤差が発生する．分析者の観点からはこの誤差はもちろん小さい方がよい．したがって，属性値を加工する際には k-匿名性の充足と，分析の誤差の低減を両立することが望ましい．

k-匿名化技術とは，ある個票データに対して，セキュリティパラメータ k および準識別子とする属性を定めたときに，k-匿名性を満たし，かつ分析結果の誤差が小さくなるように属性値に対する加工方法を決定するアルゴリズムの総称である．

本節では，k-匿名化技術について，基本となる属性値の加工方法である“一般化”と，一般化を用いた代表的なアルゴリズムである Mondrian アルゴリズム（LeFevre et al. 2006）について紹介しよう．

3.2.1 属性値の一般化

一般化は k-匿名化で最も使用される属性値の加工方法であり，ある属性値を，その属性値を含むより上位の範囲やカテゴリに変更する加工方法である．

年齢などの数値属性の場合は図 4 の年齢の列のように，同じ一般化をしたいレコード集合の属性値をすべて含む範囲に変更する．例えば 28 歳と 30 歳を持つレコード集合を一般化したい場合，28-30 歳とする．

職業などのカテゴリ値をとる属性の場合は図 5 のように汎化木と呼ばれる木構造を定義し，レコード集合の属性値に対応するすべての葉ノードに共通する上位ノードを探し，そのノードが対応する属性値に変更することで一般化を行う．例えば職業について“大学教授”，“高校教師”を持つレコード集合を一般化する場合は職業列を“教育者”に変更する．

このような加工を実施することで属性値がばらばらのレコードが同じ値を持つようにするのである．

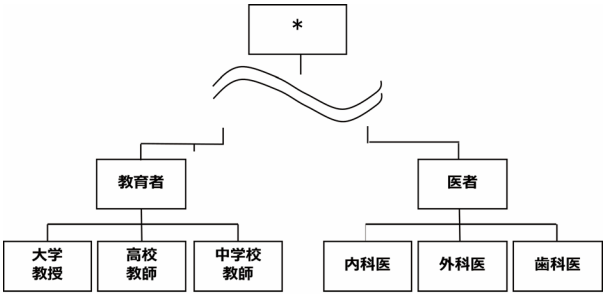

図 5 汎化木の例

3.2.2 アルゴリズム

Mondrian は個票データを，k 個以上のレコードを持つグループ集合に分割し，グループ単位で属性値を一般化することで k-匿名化を行うアルゴリズムである．

Mondrian は以下のようなステップで実行される．最初はすべてのレコードをひとつのグループとし，このグループに対して，準識別子と分割点となる属性値（例えば中央値）を選択し，分割点を境目としてグループを 2 分割する．そして，分割された新しいグループそれぞれに対して同様の分割を再帰的に繰り返していく．グループをこれ以上分割できない，つまり 2k 個以上のレコードがグループに含まれない場合，再帰処理を終了する．このような処理により，同じ値へと再符号化を行うレコード群をグループ分けする技術である． 図 6 は 2 つの数値をとる準識別子（例えば年齢と郵便番号）を持つデータへ k=2 として Mondrian アルゴリズムを適用したときの動作イメージである．

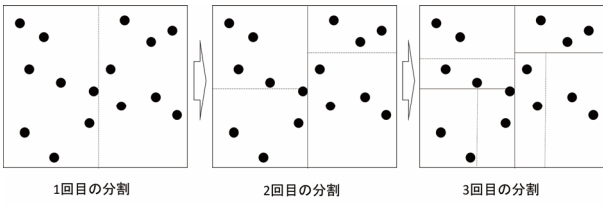

図 6 Mondorian アルゴリズムの動作イメージ

グループに分割した後は，すべてのグループに対して，含まれるすべてのレコードの属性値を一般化することで k-匿名化を完了する．

Mondrian アルゴリズムはなるべく属性値が近くなるように k 個のレコードをグループ化することで，一般化される範囲を小さくし，k-匿名性を充足しつつ分析に与える誤差を低減している．また，分割する準識別子の選択や分割する境目となる属性値の決定方法を変えることで，用途に適した k-匿名化をできる柔軟な方式である．

3.3　k-匿名化のサービス適用における注意点

3.3.1　準識別子の決定

準識別子は k-匿名化を実行するサービス事業者が自分で決定する必要がある．しかし，どの属性を準識別子と定めるのが適切なのかは，それぞれの個票データの特性やデータ分析者の目的によって異なる．また，準識別子を適切に設定しなかった場合，匿名化データに個人を識別できる情報が残ってしまい，個人の識別を容易に行えてしまう可能性がある．

準識別子選別のひとつの方法として「観測可能性」を基準として選別する方法が挙げられる．観測可能性とは，ある個人の知人であれば知っているかどうかや，過去に公表されているデータやウェブ上のデータから知ることができるかといった形で他者から知られる可能性であり，これが高い属性を準識別子とするのである．英国の権利保護機関 Information Commissioner's Office（ICO）が発行した文書（ICO 2012）では提供する個票データに対するリスク評価として，過去に開示したデータやウェブ上で得られる公開情報を用いる例を挙げている．

3.3.2　属性推定攻撃への耐性とℓ-多様性

k-匿名性は個票データから識別子，準識別子によって個人を特定できないことを表す指標であった．しかしながら，個人が特定できないだけでプライバシーは保護できていると言えるのであろうか？例えば，同じ準識別子を持つk人が同じセンシティブ属性を持っていた場合，k人が識別できなくともセンシティブ属性の値は推定できてしまう．Aさんがどのレコードに該当するか分からなくても，Aさんの病気が推定できてしまうことはプライバシー上問題であろう．このような，属性推定攻撃と呼ばれる攻撃に弱いのがk-匿名性の欠点となる（図 7）．

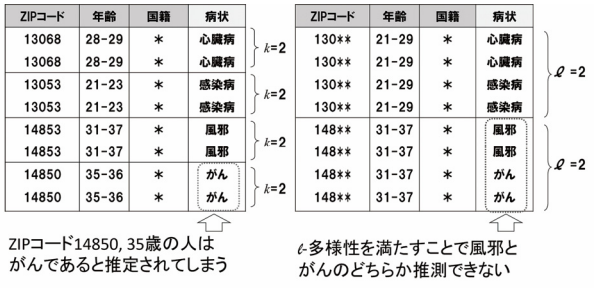

図 7 属性推定と l-多様性

この欠点に対応した指標としてℓ-多様性（Machanavajjhala et al. 2007）が提案されている．ℓ-多様性とは，k-匿名性を保った上で，さらなるセキュリティパラメータℓが与えられたとき，「同じ準識別子を持つレコードに紐づくセンシティブ属性が少なくともℓ種類以上あること」を求める性質である．この性質を満たすことにより，少なくとも個人のセンシティブ属性はℓ種類のうちのいずれかに該当する，ということまでしか推論ができないため，属性推定攻撃にある程度対応できる指標であると言える．

これらの点に注意を払うことを嫌う場合，安全性指標として k-匿名性ではなく差分プライバシー（Dwork 2006）を使用する選択肢もある．本稿では説明を割愛するが，文献（寺田 2019）などに詳しい．

4. マルチパーティ計算技術

マルチパーティ計算技術とは複数の参加者（パーティ）が，秘密の入力を秘匿したまま所定の計算を実行する技術の総称である．マルチパーティ計算技術は秘密計算技術とも呼ばれる．本章ではマルチパーティ計算技術のうち，文献（Araki et al. 2016）で提案されている秘密分散技術を用いた 3 者秘密計算技術について紹介する．これは秘密を分散させたまま，それぞれが部分的な計算をすることによって，最終的に計算結果のみを得ることを目標としている方式である．

4.1　秘密分散法

秘密分散とは，秘密にしたい情報をシェアと呼ばれる複数の情報に分割することでデータを秘匿する技術である．そして，事前に規定した k 個のシェアがそろった場合，元の情報を復元できるという性質を持つ．作成するシェアの数を n，復元に必要なシェアの数を k としたときに，(k, n)-秘密分散法などと表記される．共通鍵暗号や公開鍵暗号といった暗号技術では暗号文と鍵が対になっているが，これは，暗号文，鍵をそれぞれひとつずつのシェアとしてみたとき，(2, 2)-秘密分散になっているとみなすこともできる．

本節では，(n, n)-加法型秘密分散法と，それに基づいた複製型秘密分散法について紹介する．複製型秘密分散法の形で秘密を分散することにより，文献（Araki et al. 2016）では高速なマルチパーティ計算を実現している．

4.1.1　加法型秘密分散法

加法型秘密分散法では，秘密にしたい情報 m について， n-1 個のシェア$(s_1, s_2, \ldots, s_{n-1})$をランダムに生成し，式（1）を満たすよう，$s_n$を計算する．ここで q は m がとりうる値に対して m<q となるように最大値を抑えるための値である．m が符号なし 64bit 整数の場合，$q = 2^{64}$とする．

$$m = (s_1 + s_2 +, \dots, +s_n)\ mod\ q \qquad (1)$$

このように生成したシェア$(s_1, s_2, \dots, s_n)$のうち，どの n-1 個をみても，m は復元できない．元の値を復元したい場合には式（1）に従い，すべてのシェアを合計し，q で剰余をとることで得ることができる．

例えば，$q = 2^{64}$の下で 100 という値を 3 つのシェアに分散することを考えてみよう．図 8 に示すように 15, 25, 60 の 3 つの値に分けて 3 人に分散したとき，3 人のうち 2 人が結託したとしても，残りのシェアを推定することは不可能であり，元の値である 100 を復元することはできない．このため，3 人が協力しなければ元の値は復元されることなく安全に分散されたと言える．

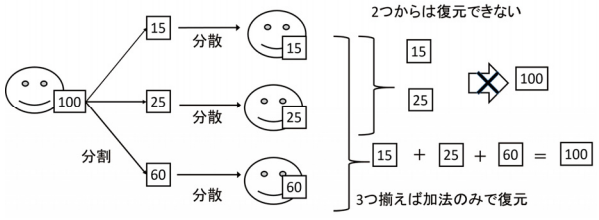

図 8 (3, 3)-加法型秘密分散法

4.1.2 複製型秘密分散法

複製型秘密分散法(Araki et al. 2016)とは，(3, 3)-加法型秘密分散法により生成した 3 つのシェアのうち, 2 つ組をシェアとすることで(2, 3)-秘密分散法を構成する秘密分散法である．つまり，加法型秘密分散により生成されたシェア(s_1, s_2, s_3)を$((s_1, s_2), (s_2, s_3), (s_3, s_1))$という形に変更するのである(図 9)．3 つのうち 2 つのシェア, 例えば$((s_1, s_2), (s_2, s_3))$がそろえば元の値を復元できるが，ひとつのシェアでは復元できないため，(2, 3)-秘密分散法となっている．

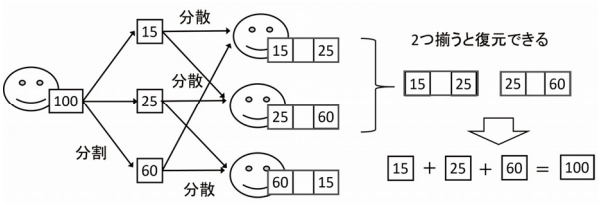

図 9 (2, 3)-複製型秘密分散法

4.2 複製型秘密分散法に基づく3者秘密計算

秘密分散を用いたマルチパーティ計算技術では，計算に用いる値それぞれに対して秘密分散法に従ってシェアを生成する．そして，それらを計算実行するパーティに分散し，そのシェアを用いながらパーティ内での計算とパーティ間での通信を繰り返すことで，元の値や計算途中の値をそれぞれのパーティに明かすことなく，所望の計算処理を実行する技術である．(Araki et al. 2016)では(2, 3)-複製型秘密分散法で作成した 3 つのシェアを 3 つのパーティで保持して計算を実行する 3 者秘密計算が提案されている（図 10）．この方式は，参加者を 3 人に限定することにより，高速なマルチパーティ計算を実現している．

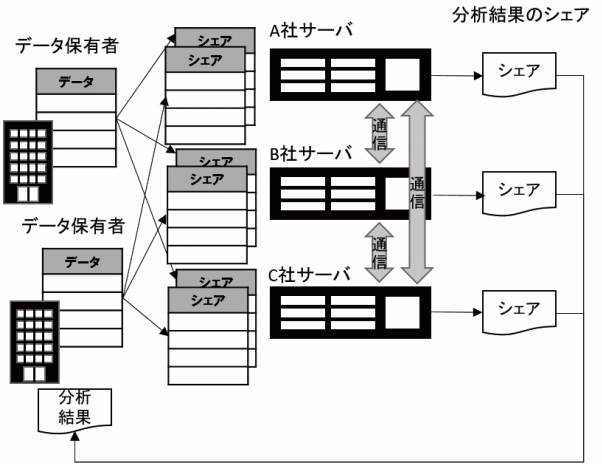

図10 3者秘密計算のシステムイメージ

4.2.1 基本演算

では，どのような計算処理が可能なのであろうか？本節では基本的な演算である加法と乗法について，その計算方法を紹介しよう．加法と乗算を組み合わせることで任意の計算処理が可能になる．

ここでは 2 つの秘密である整数 a, b が複製型秘密分散によって 3 つのパーティに分散されているものとする．i 番目のパーティが所持するシェアを(s_i, s'_i) (i=1, 2, 3) と表記する．

ここで，$a = (x_1 + x_2 + x_3)\ mod\ q$, $b = (y_1 + y_2 + y_3)\ mod\ q$ という加法型秘密分散に対して$s_i = (x_{i+1}, x_{i+2}), s'_i = (y_{i+1}, y_{i+2})$という複製型秘密分散のシェアが構成されているとしよう（添え字が 4, 5 となる場合はそれぞれ 1, 2 と読み替えてほしい）．

加法：複製型秘密分散法の加法は非常に単純であり，各パーティが所持しているシェアをローカルに計算するだけでよい．つまり，パーティ i は$z_{i+1} = x_{i+1} + y_{i+1}\ mod\ q$, $z_{i+2} = x_{i+2} + y_{i+2}$ mod q を求めるだけである．

加法で求めたいのは$(a + b)$ mod $q = (x_1 + x_2 + x_3) + (y_1 + y_2 + y_3)\ mod\ q$ である．式(2)により，新しく得られたz_1, z_2, z_3が a+b の加法型秘密分散のシェアとなっており, パーティ i が持つ(z_{i+1}, z_{i+2})は a+b の複製型秘密分散のシェアとなっていることが分かるだ

ろう．加算の結果を知りたいときは，2 つのパーティがそれぞれのシェアを持ちより $z_1+z_2+z_3$ を計算すればよい．

$$\begin{aligned}&(z_1+z_2+z_3)\ mod\ q\\&\quad=(x_1+y_1)+\ (x_2+y_2)+(x_3+y_3)\ mod\ q\\&\quad=\ (x_1+x_2+x_3)+\ (y_1+y_2+y_3)\ mod\ q\\&\quad=(a+b)\ mod\ q\end{aligned}\tag{2}$$

乗法：乗法は加法と比較して少々複雑になる．乗法にて行いたいことは，式(3)に対応する複製型秘密分散のシェアを各パーティで持つことである．

$$\begin{aligned}\mathrm{a}\cdot\mathrm{b}&=(x_1+x_2+x_3)(y_1+y_2+y_3)\ mod\ q\\&=x_1y_1+x_1y_2+x_1y_3\\&\quad+x_2y_1+x_2y_2+x_2y_3\\&\quad+x_3y_1+x_3y_2+x_3y_3\ mod\ q\end{aligned}\tag{3}$$

実は，各パーティが持つシェアをローカルに計算するだけで，式(3)におけるすべての項をそろえることが可能だと分かるであろうか？具体的には各パーティが $x_{i+1}y_{i+1}$，$x_{i+1}y_{i+2}$，$x_{i+2}y_{i+1}$ をローカルに計算すればすべての項がそろうのである．$z_i=x_{i+1}y_{i+1}+x_{i+1}y_{i+2}+x_{i+2}y_{i+1}$ とすれば，$z_1+z_2+z_3=a\cdot b$ を満たすことが分かるであろう．つまり，z_1,z_2,z_3 は $a\cdot b$ の加法型秘密分散のシェアを構成していることになる．

複製型秘密分散のシェアを構成するには加法型秘密分散のシェアを 2 つずつ持つ必要がある．このため，複製型秘密分散の乗法ではパーティ 1 から 2，2 から 3，3 から 1 といったふうに自分が持つシェアを共有する．したがって，各パーティの間で通信が 1 度ずつ発生することになる．

実際には z_i をそのまま送信すると元の a, b に関する情報が漏洩する可能性がある．このためセキュアに $(r_1+r_2+r_3)\ mod\ p=0$ を満たす乱数 r_i をそれぞれのパーティで発生させ，$z'_i=z_i+r_i$ とした値をシェアとして用い，隣のパーティと共有する．セキュアな乱数の発生のさせ方などの詳細は原論文を参照されたい．

ここで紹介した加法，乗法の結果は再びシェアの形でパーティ 1, 2, 3 に分散されていることに注目してほしい．このことは，演算後にさらに加法や乗法を実施できることを意味している．このように，加法や乗法を繰り返すことによって任意の計算ができるのである．

4.3 マルチパーティ計算のサービス適用における注意点

マルチパーティ計算技術の実用上の注意点として，まず，実行速度が通常のコンピュータ上での演算と比較して非常に遅いことが挙げられる．今回紹介した方式はその中でも効率のよい方式ではあるが，それでも乗法のようにパーティ間の通信が必要となる演算でそれが顕著となる．現時点のコンピューティングリソースでは大規模な機械学習といった重い処理は実現が難しいかもしれない．できる範囲の処理を適切に見極める必要があるだろう．

また，本稿で紹介した 3 者秘密計算では，3 つのパーティがそれぞれ結託しないことが安全性を担保するために必要である．2 章で述べたパーソナルデータの委託分析に 3 者秘密計算を適用する際には，3 つのパーティを別の事業者が管理するような工夫が求められる．その上で，契約により各事業者の結託を禁止するような法的拘束力が必要となるケースもあるだろう．

さらに，例で見たようにデータ a, b の和と積を計算してしまうと，マルチパーティ計算をしていても，和と積の結果から a, b の組を特定できてしまう．同じデータに対して不用意な計算をしてしまわないよう，具体的にどのような計算の実施を許可するかの制御も別途検討する必要がある．

5. おわりに

本稿ではパーソナルデータの保護と活用の両立によってサービスをより豊かなものにする一助となればと思い，パーソナルデータ秘匿化技術とそれらの技術が適するサービスの形態について紹介した．

これらの技術を広く現実のサービスに活用していくには，より社会受容性を高めていく必要がある．例えば，k-匿名化技術を適用する際に準識別子の選択を正しく行わなければ匿名化データから容易に個人を識別・特定できてしまうだろう．また，秘密計算技術も出力された分析結果から入力データを推測できてしまう可能性がある．こういった問題を起こさないための条件や制度をどうするかといった，運用面を改善していくことが技術の社会受容性を高めるひとつの方法であろう．

このためには，技術の実ユースケースを積み重ね，それらの中からベストプラクティスを導出することで，社会的なルールへ昇華していくことが重要である．このような未来の実現には，パーソナルデータ秘匿化技術の研究者だけでは不十分であろう．サービス事業者やサービス学研究者の今後の貢献を期待したい．

◆ 参考文献 ◆

Araki, T. Furukawa, J, Lindell, Y. Nof, A. Ohara, K. (2016). High-Throughput Semi-Honest Secure Three-Party Computation with an Honest Majority. CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 805-817.

Boston Univ (2019). http://www.bu.edu/articles/2019/secure-multiparty-computation/

CNIL (2019). https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc

Dwork, C. (2006). Differential privacy. Proc.33rd Intl.Conf. Automata, Languages and Programming, Volume Part II, Vol.4052 of LNCS, 1-12

EU (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

Information Commissioner's Office (ICO) (2012). Anonymisation: managing data protection risk code of practice.

Information Commissioner's Office (ICO) (2019). https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2019/07/ico-announces-intention-to-fine-british-airways/

LeFevre, K. Dewitt, D,J, Ramakrishnan, R. (2006). Mondrian Multidimensional k-Anonymity. 22nd International Conference on Data Engineering, 25-35.

Machanavajjhala, A. Kifer, D. Gehrke, J. Venkitasubramaniam, M (2007). l-diversity: Privacy beyond k-anonymity. ACM TKDD, 1 (1), 3.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570.

寺田雅之 (2019). 差分プライバシーとは何か. システム／制御／情報, 63 (2), 58-63.

◇ 著者紹介 ◇

古川　諒

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

佐古　和恵

京都大学理学部（数学）卒業後, NEC にて特別技術主幹を経て早稲田大学基幹理工学部教授. 暗号プロトコル技術を用いたセキュリティ・プライバシ保護・公平性保証の研究に従事. 博士（工学）.