URL https://journal.ntt.co.jp/article/23744
DOI https://doi.org/10.60249/23113101

企業横断の統計的なデータ活用による社会課題解決を実現する秘匿クロス統計技術

さまざまな社会課題の解決には，各企業の保有するデータを企業横断で活用し，社会で発生する事象を俯瞰的にとらえることが有効です．一方，国内外でデータ提供者のプライバシ保護に関する新たなルールづくりも進んでおり，個人のプライバシを保護しつつ，データ活用を行うことが重要になっています．こうした背景を踏まえて，NTTドコモではプライバシ情報を保護したうえで，企業横断で統計情報を作成する秘匿クロス統計技術の研究開発を進めています．秘匿クロス統計技術により，企業を横断したデータの統計的活用を通じて，従来得られなかった新たな視点で事象をとらえ，効果的な課題解決につながることが期待されます．

はじめに

データに基づく意思決定は広く社会に浸透し，さまざまな社会課題の解決においてデータ活用の重要性が増しています．データを用いて事象を正しくとらえることで，より効率的に問題解決を図ることが期待されます．しかし，ある特定の問題に注目し，その問題を解決しようと考えた場合，単独の企業が保有するデータだけでは生じている事象を一面的にしかとらえられない可能性があります．そこで，複数の異なる企業に蓄積されたデータを複数企業で活用することで，複合的な観点や俯瞰的な視点から社会で発生する事象をとらえることが考えられます[(1)]．

一方で，国内外でデータ提供者のプライバシ保護に関する新たなルールづくりも進んでいます（例えば，EUではGDPR：General Data Protection Regulation[(2)]*1や日本では改正個人情報保護法[(3)]など）．これらのルールの遵守はデータ活用を行う企業の責務となります．また，このような法令遵守のみならず，データ主体である個人が安心するかたちでプライバシ情報を保護し，データ活用を行うことが重要となっています．関連法令を遵守し，かつ個人のプライバシ情報を保護したうえで，データを活用する方法として，個人を特定できないようにデータを統計化した「統計情報」への作成および活用（統計的データ活用）が考えられます．統計情報とは，複数人の集団の情報から共通要素を集計するなどで得られる個人との対応関係が排斥された情報であり，集団の傾向や性質のみを示します．企業単独の統計的データ活用は多く行われていますが（例えば，文献⑷など），企業横断でそれを行うことは容易ではありません．その理由としては，企業横断の統計的データ活用は，企業間でデータの受け渡しが発生する過程においても，関連法令を遵守し，かつプライバシ情報を明かさずに，統計情報を作成することが必要となるからです．

NTTドコモは，NTT社会情報研究所と協力し，企業が保有するデータを横断した統計的データ活用を実現する技術として秘匿クロス統計技術（**図**）を開発しました．秘匿クロス統計技術は各社が保有するデータ（例えば，NTTドコモのデータとパートナー

＊ 本記事は「NTT DOCOMOテクニカル・ジャーナル」（Vol.31 No.1, 2023年4月）に掲載された内容を編集したものです．

＊1 一般データ保護規則（GDPR）：EU加盟国および欧州経済領域で適用される個人情報の取扱いに関する保護規則．個人情報の取得や移動にも適用されます．

図　秘匿クロス統計技術

企業のデータ）を，各社において個人を識別できない状態（個人情報ではない状態）に加工したうえで，データを相互に明かすことなく，すなわち，一連の処理を人の目に触れることなく機械が行うことを技術的に保証して，安全な統計情報を作成する技術です．ここでは企業横断での統計的データ活用における要件を整理し，要件を達成する方針を述べ，秘匿クロス統計技術の概要を解説します．

秘匿クロス統計技術と社会課題解決に向けた活用

■企業横断の統計的データ活用における要件

各社が保有するデータから統計情報を作成するためには，通常ならば，少なくともいずれか一方の企業から，もう一方の企業へデータを提供し，集計処理を行うことが必要です．その際に提供先の企業が受領したデータの中身を確認できる場合，提供先の企業（もしくは第三者）に対して入力データが明かされてしまいます．また，データ連携後に出力されるデータに個人との対応関係が排斥されていない場合，出力データのプライバシ情報が保護されません．

さらに，いかに安心・安全な出力データが得られたとしても，そのデータに有用性がなければ価値はありません．また，その有用性は，企業が保有するデータを横断して作成した以上，単独企業では分かり得なかったデータ活用の価値であることが望ましいため，安心・安全な統計情報を作成するだけではなく，その有用性についても評価する必要があります．企業横断の統計的データ活用における，満たすべき要件を下記に示します．

① データ連携前に個人を識別できないデータに加工するとともに，データ連携中に自社のデータが他社に明かされないこと
② データ連携後の出力データにおけるプライバシ情報が保護されること
③ 企業横断で作成した統計情報から，単独企業では得られない価値が創出されること

■安全性要件を満たすアプローチ

下記の⑴～⑷の処理を適切に組み合わせることで，要件①と②を満たすことができます．

⑴ 非識別化処理

データ連携前に個人を識別できないデータに加工するため，代表的な処理として，鍵付ハッシュ関数*2を採用しました．鍵付ハッシュ関数ではデータにソルト*3を付与したうえで，ハッシュ値（非識別化ハッシュ）を計算すること（ハッシュ化*4）ができます．ハッシュ化を行った後に，ソルトを破棄することで，不可逆に非識別化ハッシュへ変換します．

⑵ 準同型暗号技術を用いた集計処理

データ連携中に自社のデータが他社に明かされないために，準同型暗号技術[5]を採用しました．これはデータを暗号化したまま，計算処理が可能な技術です．この技術を応用することで，データを暗号化したまま，企業間でデータの集計などの演算処理が可能になります．準同型暗号技術は，許可されたもののみが情報にアクセスできるという機密性を保証し，これが相互にデータを明かさずに集計処理ができるため，出力データを作成する過程のプライバシ保証の課題への対策となります．

⑶ 差分プライバシに基づくノイズ付加による秘匿処理

企業間の統計的データ活用の実現には，出力データの作成過程のみならず，出力データ自体のプライバシ保証が必要となります．そのため，出力データのプライバシ保証の課題への対策には，差分プライバシ*5に基づくノイズ付加*6を実施します．差分プライバシとは，特定の背景知識や攻撃能力を持つ攻撃者に対しても安全性を保証できることを目的として作成された，プライバシ保護の強度を定量的に測る指標です[6]．従来の指標は，特定の攻撃者や前提条件に対する安全性の保証度合いを示すものでしたが，差分プライバシは，汎用的な安全性を目的とした指標です．各社のデータを準同型暗号技術で暗号化し，暗号状態のまま集計処理と差分プライバシに基づくノイズ付加を行うことで，出力データのプライバシ保証の課題への対策となります．仮に，差分プライバシに基づくノイズ付加を平文の状態で行った場合には，ノイズ付加前の集計結果を知得できてしまうため，入力から出力までの一連のデータ処理をすべて暗号化された状態で実施します．

⑷ データ処理の隔離実行環境への実装

準同型暗号と差分プライバシに基づくノイズ付加が正しく行われなかった場合（例えば，ソフトウェアが改ざんされるなど）には，「出力データを作成する過程のプライバシ」と「出力データのプ

---

*2 鍵付ハッシュ関数：出力された文字列からは，入力された文字列を得ることが不可能という特性を持つ一方向関数の一種．任意の長さの文字列を固定長の文字列（ハッシュ値）に変換する関数であり，同一の入力に対しては，対応する同一の文字列が出力される特性を持ちます．
*3 ソルト：データをハッシュ化する際に，鍵付ハッシュ関数の入力に加えるランダムなデータ．
*4 ハッシュ化：鍵付ハッシュ関数により元データからハッシュ値を計算すること．なお，ハッシュ化後に，ハッシュ化する際に用いたソルトを破棄するため，ハッシュ値から元のデータを算出することは不可能です．
*5 差分プライバシ：任意の背景知識や攻撃能力を持つ攻撃者に対しても安全性を保証できることを目的として作成されたプライバシ保護の強度を定量的に測る指標．なお，米国国勢調査においても，「差分プライバシ」を用いた保護手法が採用されています．
*6 ノイズ付加：出力データからプライバシ情報を保護するため，集計表に対して，乱数を付与すること．

ライバシ」のいずれのプライバシも保証されない可能性があります．したがって，ソフトウェアが改ざんされていないという性質である完全性を保証することで，期待する処理が正しく実施されることが技術的に保証されます．具体的には，準同型暗号化と差分プライバシ基準に基づくノイズ付加の一連のデータ処理を隔離実行環境に実装する方針としました．隔離実行環境とは，データを隔離された信頼できる領域に配置し，データ処理をその領域内に完結して実行する技術です．特にハードウェアに基づいて隔離された領域内でデータ処理を実行することで，隔離実行環境の中で動作しているマシンイメージやアプリケーションの完全性を保証します．

■秘匿クロス統計技術

NTTドコモでは，前述の方針を踏まえて，準同型暗号技術，差分プライバシに基づくノイズ付加技術，隔離実行環境を適切に組み合わせ，企業横断の統計的データ活用を実現する秘匿クロス統計技術を開発しました．秘匿クロス統計技術では隔離実行環境で①非識別化処理（個人を識別できない状態に加工する処理），②集計処理，③秘匿処理を行うことで，安全な統計情報を作成します．①非識別化処理は各社が保有するIDを不可逆変換し，個人を識別できない状態に加工します．具体的には，隔離実行環境内でIDにソルトを付与し，一方向関数によって，非識別化ハッシュを得た後，ソルトの破棄を技術的に保証し，非識別化ハッシュを各社のそれぞれの暗号鍵で暗号化を実施します．②集計処理と③秘匿処理は，隔離実行環境内で準同型暗号技術と差分プライバシに基づくノイズ付加技術を組み合わせた処理であり，相互に非識別化ハッシュを明かすことなく，プライバシ情報が保護された安全な統計情報を得ることができます．

社会課題の解決に向けた本技術の活用

企業横断で作成した統計情報から，単独企業では得られない価値が創出されることという要件を満たしているかどうかを検証するために，日本航空株式会社（日本航空），株式会社ジャルカード（JALカード）と，スムーズな航空利用の実現を通じて顧客体験価値向上と社会課題の解決に取り組む実証実験を実施しました[7]．

この実証実験では，日本航空が保有する国内線航空券の予約データに関する情報に，NTTドコモが保有する携帯電話ネットワークの運用データ*7（携帯電話の位置データを含む）を，本技術を用いて各社が保有するデータを各社において，個人を識別できない状態（個人情報ではない状態）に加工したうえで組み合わせることで，航空機搭乗前の空港内外両方での移動状況に関する人口統計情報を作成しました．この人口統計情報を活用することで，単独企業では得られなかった新たな視点を得ることができ，秘匿クロス統計技術の有用性を確認することができました．さらに，顧客体験価値向上と社会課題の解決に取り組む実証実験の第2弾として，日本航空，JALカード，株式会社北海道エアシステムと，北海道内の移動ニーズを把握する実証実験を実施しています[8]．北海道の道東エリアを対象として，お客さまの移動に関する人口統計情報から，空港を中心とした移動状況を把握し，単独企業では得られなかった利便性の高い交通手段の提供および道東エリアの人流創出につながる知見の獲得をめざしています．

おわりに

ここでは，企業横断の統計的データ活用が満たすべき要件とそれを満たす秘匿クロス統計技術について解説しました．今後は，秘匿クロス統計技術によりデータの活用と保護を両立させ，さまざまな社会課題の解決につなげることで，社会と産業のさらなる発展に結びつけていきます．

■参考文献

(1) 経済協力開発機構（OECD）：“OECDビックデータ白書　―データ駆動型イノベーションが拓く未来社会，”明石書店，2018.
(2) https://eur-lex.europa.eu/eli/reg/2016/679/oj
(3) https://elaws.e-gov.go.jp/document?lawid=415AC0000000057
(4) https://mobaku.jp/
(5) 佐久間・陸：“⑤準同型暗号を用いた秘密計算技術と実用化に向けた活動，”情報処理，Vol.59，No.10,pp.898-903，2018.
(6) 寺田：“差分プライバシーとは何か，”システム／制御／情報，Vol.63，No.2，pp.58-63,2019.
(7) https://www.docomo.ne.jp/binary/pdf/info/news_release/topics_221020_00.pdf
(8) https://www.docomo.ne.jp/binary/pdf/info/news_release/topics_230822_00.pdf

◆問い合わせ先

NTTドコモ
R&D戦略部
E-mail　dtj@nttdocomo.com

*7 携帯電話ネットワークの運用データ：電気通信サービスを提供する過程で発生するデータの総称．電気通信サービスを提供する過程で発生するデータの総称であり，モバイル空間統計でも利用されています．運用データは，お客さまがご利用の携帯電話の位置データおよびお客さまの属性データを含むものであり，それぞれの定義については下記のモバイル空間統計ガイドラインを参照願います．
https://www.docomo.ne.jp/corporate/disclosure/mobile_spatial_statistics/guideline/