

## NTT持株会社ニュースリリース

2014年2月7日

## ビッグデータ時代における新たなパーソナルデータ匿名化システムを開発 ～高度にプライバシー保護したままに、データの利用価値を高いままとする～

日本電信電話株式会社(本社:東京都千代田区、代表取締役社長:鶴浦博夫、以下「NTT」)は、ビッグデータ分析に用いられる個人に関する情報(パーソナルデータ)を、高度にプライバシーを保護したまま、分析データ(ビッグデータ)の利用価値を高いままに加工することが可能な匿名化システムを開発しました。

開発したシステムは、パーソナルデータから個人が直接特定できる情報を取り除いた上で、さらにそのデータから誰か一人に絞り込めないようにデータを加工することで高度なプライバシー保護対策をします。

加工方法は、匿名性の代表的な指標である「k-匿名性」(※1)を満たす匿名化を実施し、従来から知られている希少な人のデータを取り除く「削除」、項目の値をより粗くする「一般化」に加え、NTTが独自に開発した手法である「Pk-匿名化」をパーソナルデータの種類や分析目的にあわせて柔軟に選択できるため、従来困難とされていた高度なプライバシー保護対策を行った場合の利用価値の低下を実用的なレベルで押さえた分析用データを作成することが可能になりました。

本システムをビッグデータ分析のプロセスに取り入れることにより、高度なプライバシー保護対策と安全管理措置を施した分析業務が可能になります。

### 1. 開発の背景

個人情報の取り扱いのルールは、個人情報保護法の法体系によって定められています。しかしビッグデータの時代になり、購買履歴や移動履歴といったパーソナルデータを分析する際のプライバシーの保護が大きな課題となっています。

NTTではこれまで、プライバシー保護と安全管理措置を実現するハイエンドソリューションとして、暗号化したままデータ処理を行う秘密分散・秘密計算技術(※2)の研究開発を行ってきました。これらの技術群はデータの正確さを全く損なわないまま、その秘匿性を極限まで追求するものです。

一方で、個人情報をいわゆる「匿名化データ」に加工して分析する際のプライバシー侵害のリスクを低くしたいというニーズも存在しています。いわゆる匿名化において、例えば、氏名と年齢と年収から構成される古典的なパーソナルデータを分析する場合、そのデータから氏名を消去すればプライバシー上の問題は少ないとされていました。しかしビッグデータ時代においては、個人に関する多数の項目や長期間にわたる履歴がパーソナルデータに含まれるため、たとえ個人が直接特定できる氏名や住所を消去しても、残りの項目と他の情報を突き合わせることで個人が特定されてしまうリスクが指摘されています。

このリスクに対して、データをひとりひとりに識別することを困難とするデータ加工の指標「k-匿名性」が注目されています。k-匿名性をパーソナルデータに完全に適用すると、そのデータからは誰の情報なのかk人未満に絞り込むことができません。しかし、NTT研究所では、k-匿名性を単純にビッグデータに適用すると以下のような困難な点があるとの観点にたち、これらの課題を解決するための研究開発を行ってきました。

1. 匿名性を高めるためにはデータを粗く加工する必要があるが、データを粗くすると元データとかけ離れ利用価値が低下する(匿名性と有用性のトレードオフ)
2. 匿名性を高めるためにはデータのより多くの項目を加工対象にする必要があるが、データの加工対象の項目を増やすと利用価値が低下する(ビッグデータ匿名化のジレンマ)

### 2. 開発したシステム

今回NTTセキュアプラットフォーム研究所が開発した匿名化システムは、パーソナルデータを保有する事業者向けのソフトウェアで、匿名性と有用性のトレードオフをバランスさせk-匿名性を満たした上でデータの有用性が損なわれにくい匿名化データを得ることができる特徴を持ちます。このため、自社内外で分析を行う際にこの匿名化データを用いることで、プライバシー保護の高度な対策を行ったパーソナルデータ分析を行うことができます。事業者は、保有するパーソナルデータを匿名化システムに入力し、(1) 匿名化処理方法の選択、(2) 分析に用いるパーソナルデータの項目の選択、(3) 匿名性のパラメータ(kの値)の決定をすると、k-匿名性を満たしかつデータの有用性の高い「匿名化データ」が出力されます(図1□)。

本システムはk-匿名性を確保する匿名化処理として、「削除」、「一般化」、NTTが独自に開発した「Pk-匿名化」(図2□)を備え、データの種類や分析目的に応じて使い分けことができます。「削除」は頻出データの分析に、「一般化」は全体傾向の把握に、「Pk-匿名化」は「長いデータ」の分析に有用です。特に、多項目を持つ「横に長い」ビッグデータの匿名化を行うときは、ビッグデータ匿名化のジレンマを回避するために、分析項目を選択して「横に短い」データにして加工する「オーダーメイド匿名化」を推奨しています(図3□)。Pk-匿名化を用いれば、分析の目的に合わせて異なる項目の組みあわせでオーダーメイド匿名化を繰り返す場合でも、全体を通じてk-匿名性が損なわれません。

### 3. 今後の展開

今回開発した匿名化システムは、NTTグループを通じて今後半年以内を目途に、パーソナルデータを保有する事業者様向けに、ビッグデータ分析支援のソリューションの一環として提供する予定です。

なお、NTTが独自に開発した「Pk-匿名化」技術に関する詳細は以下の通りです。

### 4. 開発した技術の特徴

NTTが独自に開発した「Pk-匿名化」は「ランダム化」の一種で、個々のデータを確率的に変化させる処理と「ベイズ推定」と呼ばれる機械学習の手法により元の状態を推定する処理を行い、k-匿名性を満たした利用価値の高いデータを作成します(図4□)。これまでランダム化したデータの匿名性指標は明らかになっていませんでしたが、NTTは世界で初めてk-匿名性と等価な安全性を持つランダム化法を開発し、そのことを理論的に示しました。

Pk-匿名化によって作成されるデータは、従来の削除と一般化によるものと比べ、理論的に同等のk-匿名性を持った上で、十分に実用的な分析が行えることが実験的に明らかになっています。また、Pk-匿名化は多数の項目を持つパーソナルデータを匿名化する場合に有効です。従来の手法で知られていた、繰り返しオーダーメイド匿名化をした複数の匿名化データを突き合わせることによるk-匿名性の喪失の問題に対する耐性を持つため、繰り返しオーダーメイド匿名化ができ、多項目のパーソナルデータに対してもプライバシーを保護し利用価値が高い分析用データに加工することができます。

### 用語解説

※1 k-匿名性:匿名化したデータから個人の識別が困難であることを示す安全性の代表的な指標で、“複数の項目で同じ値の組みあわせが少なくともk個存在すること”を表す。

※2 医療統計処理における秘密計算技術を世界で初めて実証

<http://www.ntt.co.jp/news/2012/1202/120214a.html>

### 別紙・参考資料

- ▶ [図1 匿名化システムの処理イメージ](#)□
- ▶ [図2 匿名化データの加工方法](#)□
- ▶ [図3 「オーダーメイド匿名化」](#)□
- ▶ [図4 Pk-匿名化の加工イメージ](#)□

### 本件に関するお問い合わせ先

■ NTTサービスイノベーション総合研究所  
広報担当  
TEL (046) 859-2032 E-mail: randd@lab.ntt.co.jp



Innovative R&D by NTT

NTTのR&D活動を「ロゴ」として表現しました

ニュースリリースに記載している情報は、発表日時点のものです。現時点では、発表日時点での情報と異なる場合がありますので、あらかじめご了承くださいとともに、ご注意をお願いいたします。

[NTT持株会社ニュースリリース インデックスへ](#)

## NTT持株会社 ニュースリリース

[▶ 最新ニュースリリース](#)

[▶ バックナンバー](#)

[▶ English is Here](#)

## NTT持株会社 ニュースリリース内検索

1997 ▼ 年 04 ▼

月 ~

2021 ▼ 年 11 ▼ 月

検索

NTTグループの情報は  
こちらからもご覧いただけます。



[▲ このページの先頭へ](#)

[▶ 更新履歴](#) [▶ サイトマップ](#) [▶ お問い合わせ](#) [▶ 著作権](#) [▶ プライバシーポリシー](#) [▶ 情報セキュリティポリシー](#) [▶ ウェブアクセシビリティポリシー](#) [▶ 個人情報保護について](#)

Copyright © 2021 日本電信電話株式会社