

2018年11月20日
国立大学法人東京大学
株式会社日立製作所
科学技術振興機構（JST）
内閣府

大規模データの匿名加工処理を高速化する技術を開発 ～ データの有用性とプライバシー保護を両立する対話的な匿名加工を可能とし、 パーソナルデータの安全な利活用を促進 ～

1. 発表者：喜連川 優（東京大学 生産技術研究所 教授）

2. 発表のポイント：

- ◆ 情報化社会の進展に伴い、個人情報を含む大規模データの活用が求められています。しかし、大規模データに対して、目的にかなう情報量を確保しながら、プライバシー保護のために匿名加工を施すには、膨大な時間がかかることが課題でした。
- ◆ 大規模データの匿名加工処理の時間を大幅に短縮し、従来の方法の100倍程度の高速性で、対話的に細かな調整をしながら繰り返しデータの抽出・加工・検証を行える技術を開発しました。
- ◆ 本技術は、ヘルスケアや都市計画、購買履歴分析などへのパーソナルデータの活用への適用が期待されます。また、2019年度中に製品に実装することが計画されています。

3. 発表概要：

東京大学 生産技術研究所（所長：岸 利治、以下、東大生研）の喜連川 優 教授の研究グループと株式会社日立製作所（執行役社長兼 CEO：東原 敏昭、以下、日立）は、内閣府 総合科学技術・イノベーション会議が主導する革新的研究開発推進プログラム（ImPACT）「社会リスクを低減する超ビッグデータプラットフォーム」（プログラム・マネージャー：原田 博司、以下、ImPACT）の支援の下、ハードウェア性能を最大限に引き出すソフトウェア実行原理である非順序型実行原理（注1）を活用し、大規模データの匿名加工処理を高速化する技術を共同で開発しました。これにより利用目的に応じたデータの有用性とプライバシー保護を両立するための対話的な匿名加工処理を可能とし、パーソナルデータの安全な利活用を促進します。

本技術は、2018年11月21日（水）にベルサール九段において開催される「ImPACT シンポジウム」（注2）にて展示、および2018年12月21日（金）、22日（土）に国立情報学研究所において開催される電子情報通信学会データ工学研究会第一種研究会・情報処理学会データベースシステム研究会合同研究会（注3）にて発表予定です。

4. 発表内容：

近年、情報化社会の進展に伴い、地方自治体や公共機関、医療機関、民間企業などが保有するさまざまなデータを有効活用した施策の立案や、新たなサービスの創出が求められています。個人の特定が危惧されるパーソナルデータの活用においても、個人情報保護法の改正により、事業者間でデータを流通させる場合は個人を識別できないように

加工した匿名加工情報（注 4）とすることが規定されるなど、活用のための環境整備が進んでいます。

データの匿名加工処理においては、例えば同じ属性を持つデータが一定数以上存在するようにデータを変換し、個人が特定される確率を低減するといった措置を行います。しかし、このような匿名加工処理の過程では情報が失われる可能性があり、有用な情報量を確保するためには、データの抽出範囲や加工単位などを細かく調整しながら、データの検証を繰り返す必要があります。そのため、データの規模や種類が増えるほど、匿名加工処理に費やす時間が膨大になるという課題がありました。

東大生研と日立は、大規模データの匿名加工処理にかかる時間を大幅に短縮し、対話的に細かな調整を行いながらデータの抽出・加工・検証を行うことを実現する技術を開発しました。

開発した技術の特長は以下の通りです。

1. 大規模データに対する匿名加工処理の高速化技術の開発

東大生研と日立は、大規模データの解析問合せにおいて高速性を発揮してきた非順序型実行原理を、新たに匿名加工処理に適用しました。

これまで匿名加工処理には専用ソフトウェアを用いるのが一般的であり、データベースから対象となるデータを抽出した後、当該専用ソフトウェアを用いて匿名加工処理を行っていたため、大規模データに対して匿名加工処理を実施するには困難が伴いました。この度、東大生研と日立は、匿名加工処理の手続きをデータベース上の演算として定義し、匿名加工処理を非順序型実行原理に基づくデータベースエンジン上で直接実行することを可能とすることで、大規模データの匿名加工処理の高速化を実現しました。

これらの研究開発により、ユーザはデータベースからの匿名加工処理を施されたデータの抽出から抽出したデータの安全性や有用性の検証までを短時間に行えるようになり、必要に応じて匿名加工処理や検証の再実行を繰り返し行うという対話的な処理を可能としました。

2. 匿名加工処理にかかる性能の検証

東大生研は、ベンチマーク用データセットを用いた実験を行い、従来の非順序型実行原理を採用しないデータベースエンジンと今回開発した技術を適用したデータベースエンジンのそれぞれにおいて、データの匿名加工処理と有用性・安全性検証にかかる時間を計測しました。この結果、従来の非順序型実行原理を採用しないデータベースエンジンに対して、今回開発した技術により匿名加工処理および有用性・安全性検証を大幅に短縮し、100倍程度に高速化することが可能であることを確認しました（注 5）。

今回開発した大規模データの匿名加工処理を高速化する技術は、これまで利活用が困難であったパーソナルデータを含む大規模データに対して、有用性を維持しつつ安全に利活用するための新技術です。本技術は、ヘルスケアや都市計画、購買履歴分析などへの適用が期待されており、日立はさらなる大規模データの利活用の拡大に向けて、2019年度中にデータベースソフトウェア製品である「Hitachi Advanced Data Binder（注 6）」に本技術を組み込み、実用化することを計画しています。

本成果は、以下のプログラム・研究開発課題によって得られました。

内閣府革新的研究開発推進プログラム (ImPACT)

プログラム・マネージャー : 原田 博司

研究開発プログラム : 「社会リスクを低減する超ビッグデータプラットフォーム」

研究開発課題 : 「超高速動的スケーラブルデータベースエンジンの基盤技術の研究開発」

「超高速動的スケーラブルデータベースエンジンの実用化技術の研究開発」

研究開発責任者 : 喜連川 優 (東京大学)

原 憲宏 (日立製作所)

研究期間 : 平成 28 年度～平成 30 年度

本研究開発課題では、一日に数億件程度、年間では数百億件程度に上る超ビッグデータを数分～数十分程度で処理可能とするため、複数ノードへのエラスティシティ (伸縮可能性) を備えた超高速動的スケーラブルデータ処理技術を確立し、毎秒 1,000 万回程度のストレージアクセス性能を備えた新たな「超高速動的スケーラブルデータベースエンジン」の実現を目指す研究開発に取り組んでいます。

■原田 博司プログラム・マネージャーのコメント■



本研究開発プログラムは、現状のビッグデータ規模を遙かに凌ぐ「超ビッグデータ」時代に向けて、広域通信ネットワークと超高速データベース処理の 2 柱を統合した基盤技術を世界に先駆けて確立するとともに、この基盤技術により、国民 (ヒト) と生産現場 (工場) の健全性維持のための課題解決に挑戦しています。超ビッグデータ時代においては、個人情報を含む大規模データの活用が求められます。このプライバシー保護のために匿名加工が必要になりますが、目的にかなう情報量を確保しながら匿名加工を施すには膨大な時間がかかります。今回の成果は本プロジェクトで開発してきた超ビッグデータ処理エンジンを匿名化処理に適用させ従来比 100 倍程度の高速化を実現したものです。この研究開発の成果によりパーソナルデータの匿名化が加速され、超ビッグデータを利用したさまざまな価値創造が行われることが期待されます。

5. 問い合わせ先：

<研究に関すること>

東京大学 生産技術研究所

特任准教授 合田 和生 (ゴウダ カズオ)

〒153-8505 東京都目黒区駒場 4-6-1

Tel : 03-5452-6594 Fax : 03-5452-6577

株式会社日立製作所 研究開発グループ

問い合わせフォーム：<https://www8.hitachi.co.jp/inquiry/hqrd/news/jp/form.jsp>

<ImPACT 事業に関すること>

内閣府 革新的研究開発推進プログラム担当室

〒100-8914 東京都千代田区永田町 1-6-1

Tel : 03-6257-1339

<ImPACT プログラム内容および PM に関すること>

科学技術振興機構 革新的研究開発推進室

〒102-0076 東京都千代田区五番町 7 K's 五番町

Tel : 03-6380-9012 Fax : 03-6380-8263

E-mail : impact@jst.go.jp

6. 用語解説：

注 1) 非順序型実行原理

喜連川 優 東京大学 生産技術研究所 教授／国立情報学研究所所長・合田 和生 東京大学 生産技術研究所 特任准教授が考案した原理で、データの要求順序とは無関係な順序に非同期的にデータを処理することにより、ハードウェアの処理性能を最大限に引き出すことを可能にする点に特徴があります。当該実行原理に基づき東大生研と日立が開発した超高速データベースエンジンは、マルチコアプロセッサならびにストレージシステムの利用効率の著しい向上を実現することにより、ビッグデータに対する検索処理の飛躍的な高速化を実現しています。

注 2) ImPACT シンポジウム

「社会リスクを低減する超ビッグデータプラットフォーム」 2018 年度シンポジウム。詳細は下記 URL をご覧ください。

https://www.jst.go.jp/impact/sympo/hharada_2018/index.html

注 3) 電子情報通信学会データ工学研究会第一種研究会・情報処理学会データベースシステム研究会合同研究会

詳細は下記 URL をご覧ください。

https://www.ieice.org/ken/form/index.php?tgs_regid=a9056446beeb5ea1b42a135969a440b024efea2458fdc495e3efd93b78c138c5&cmd=info&lang=

注 4) 匿名加工情報

特定の個人を識別することができないように匿名加工処理を施した情報。

注 5) 100 倍程度の高速化

合計 24 コアのプロセッサ、256GB の主記憶を搭載したサーバおよび 256 台の磁気ディスクドライブを搭載したストレージシステムから構成される実験環境において、従来の非順序型実行原理を採用しないデータベースエンジンと今回開発した技術を適用したデータベースエンジンを用い、ベンチマーク用データセット（約 1,000 億レコード）を格納するデータベースから特定の属性条件に基づきレコードを選択するデータ抽出処理を対象に、匿名加工処理および有用性・安全性検証の所要時間を計測し、100 倍程度の高速性を確認しました。（例えば、従来技術の場合には 20 分程度掛かる匿名加工処理について、今回開発した技術を適用することにより、10 秒程度で答えを得られることを確認しました。）

注 6) Hitachi Advanced Data Binder

内閣府の最先端研究開発支援プログラム「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」（中心研究者：喜連川優 東京大学 生産技術研究所 教授／国立情報学研究所所長）の成果を利用しています。

このニュースリリース記載の情報(製品価格、製品仕様、サービスの内容、発売日、お問い合わせ先、URL 等)は、発表日現在の情報です。予告なしに変更され、検索日と情報が異なる可能性もありますので、あらかじめご了承ください。
