

**双方向変換の深化による自律分散ビッグデータの相互運用
基盤に関する研究**
Software Foundation for Interoperability of Autonomic
Distributed Data based on Bidirectional Transformations

課題番号：17H06099

胡 振江 (HU, ZHENJIANG)

国立情報学研究所・アーキテクチャ科学研究系・特任教授



研究の概要（4行以内）

近年ビッグデータの利活用は進んでいる中、データを集めることなく自律分散的にビッグデータの効率的な分析、共有、相互活用を行なう新しいソフトウェア基盤技術が求められている。本研究は、双方向変換技術を「高信頼・大規模・高効率」の方向へ深化させるとともに、自律分散ビッグデータの統合・共有・相互運用のためのソフトウェア基盤技術を確立する。

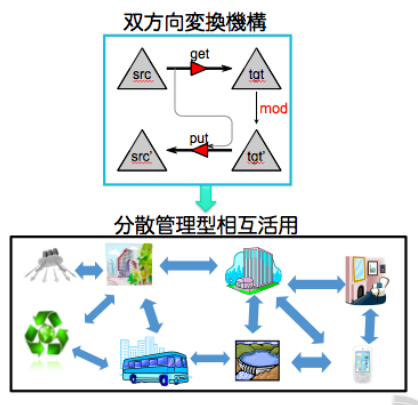
研究分野：計算基盤

キーワード：双方向変換、相互運用、自律分散データ、プログラミング言語

1. 研究開始当初の背景

近年ビッグデータの利活用は進んでいるが、データを集中管理する手法はネットワーク技術の点で厳しい局面を迎えている。その打開策として、データを集めることなく自律分散的にビッグデータの効率的な分析、共有、相互活用を行なう新しいソフトウェア基盤技術が求められている。

一方、2つのデータベース間での整合性を保証する手法として双方向変換がある。古くはデータベース分野におけるビュー更新問題として扱われてきたが、近年は新しいプログラミングモデルとデータ同期・相互運用の方法として注目を浴び、様々な双方向変換言語が提案されている。双方向変換にはデータの相互活用をはじめ多くの潜在的な応用があるが、概念実証の応用例しか発表されていないのが現状である。



2. 研究の目的

本研究は、双方向変換技術を「高信頼・大規模・高効率」の方向へ深化させるとともに、自律分散ビッグデータの統合・共有・相互運用のためのソフトウェア基盤技術を確立することを旨とする。

3. 研究の方法

本研究の目的を達成するため、まず3つの目標を設定した。「目標1: 双方向変換の深化」ではモジュール化、静的解析・自動検証、強力なデバッグ機構と学習支援により大規模・高信頼で系統的な開発を可能とする。「目標2: 双方向変換による新たな基盤の構築」では自律分散ビッグデータの相互活用のための新たな基盤: ビューパッシング計算モデルを確立し、非同期並列処理により効率性を実現する。「目標3: 双方向変換の実例への応用」では社会の実問題である、推薦システム等の2つの問題に取り組み、ビューパッシング計算モデルの有用性を示す。

そして、目標毎にチームを構成して集中的に研究に取り組んできた。研究の進め方としては、各チームでリーダーを中心に slack をベースに議論を進め(これまで投稿数は 1.5 万件に達する)、毎月 zoom によりチーム内での打ち合わせを実施し研究を推進するとともに、3チーム間で密な連携を図るため年4回程度の f2f での集中合宿を実施した。更に、各チームとも海外研究者と共同研究を実施し多数の共著論文を発表するとともに、毎年本研究プロジェクトに関する国際ワークシ

ヨップ SFDI を企画することで、海外の著名研究者からフィードバックを受けて、研究成果の完成度を高めてきた。

4. これまでの成果

双方向変換の深化: プログラミング言語の分野で研究されてきた双方向変換技術をデータベースにおける古典的な問題であるビュー更新問題に適用した。これは、ビューからソースへの変換を指定することによりその逆の変換を自動的に生成するものであり、古典的な問題に対して斬新な発想を導入するものである。関係データベースを対象として宣言的な言語によりそれを効率的に実現する手法は、最難関ジャーナルである PVLDB に採択された。また、双方向変換言語の基盤理論として、Hoare Logic に基づいて双方向変換の検証理論を構築し、その成果をプログラミング言語に関する最難関国際会議 POPL で発表した。

双方向変換によるデータ統合・相互運用基盤の構築: 双方向変換技術をもとに、連携対象の複数の DB システムの自律性を保ちながら、これらピア間でのデータ交換・更新伝搬を行うとともに、更新伝搬の大域的な一貫性を保証する機能を有する画期的なアーキテクチャである Dejima アーキテクチャを開発した。特に、データ統合の開発環境の課題に対して、膨大なデータを含むデータベース間におけるスキーマレベルのマッチングおよび値レベルの高速な類似マッチングに適用可能な技術を考案した成果は最難関ジャーナルである PVLDB、IEEE TKDE に採択され高い評価を得ている。

実例への応用: 自律分散システムの重要な応用例であるパーソナルデータ市場を実現するための基盤技術として、時系列データにおける差分プライバシーの保護の定量的評価、差分プライバシーの概念を拡張した時空間イベントプライバシーを開発し、これらの成果を纏めた 2 件の論文が最難関ジャーナルである IEEE TKDE に採択された。また、現時点で得られている研究成果を研究組織メンバー全員を含む 18 名の共著によるビジョン論文 "Dejima: Peer Data Management by Transaction for Bidirectional Update Propagation" として纏め国際会議に投稿した。日本において異なる分野の研究者が共同で集中的かつ継続的な議論を通じ英文のビジョン論文を纏めることは極めて稀である。

5. 今後の計画

今後、今までの通りに、研究目標にそって「深化」、「基盤」、「実証」という 3 つのチームで密に連携しながら研究を進めていく。深化チームは双方向・多方向変換プログラムの基礎理論を深めるとともに、その開発支援ツールの実現を完成させる。基盤チームは、

Dejima アーキテクチャに対して、技術の展開と、完成度の向上の両面で進める。深化チーム・基盤チームと連携して多くの実例を開発しその有効性を示す。

6. これまでの発表論文等 (受賞等も含む)

Van-Dang Tran, Hiroyuki Kato, Zhenjiang Hu: Programmable View Update Strategies on Relations, VLDB 2020, Tokyo, Japan, August 31 - September 4, 2020.

Yang Cao, Yonghui Xiao, Li Xiong, Liquan Bai, Masatoshi Yoshikawa, Protecting Spatiotemporal Event Privacy in Continuous Location-Based Services, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2020.

Yaoshu Wang, Chuan Xiao, Jianbin Qin, Xin Cao, Yifang Sun, Wei Wang, Makoto Onizuka: Monotonic Cardinality Estimation of Similarity Selection: A Deep Learning Approach, SIGMOD 2020, Portland, USA, June 14-19, 2020.

Jianbin Qin, Chuan Xiao: Pigeonring: A Principle for Faster Thresholded Similarity Search, VLDB 2019, Los Angeles, USA, August 26-30, 2019.

Hsiang-Shang Ko, Zhenjiang Hu: An axiomatic basis for bidirectional programming. PACMPL 2 (POPL 2018): 41:1-41:29 (2018).

Xiao He, Zhenjiang Hu: Putback-based bidirectional model transformations. ESEC/SIGSOFT FSE 2018: 434-444.

Yang Cao, Masatoshi Yoshikawa, Yonghui Xiao, Li Xiong: Quantifying Differential Privacy in Continuous Data Release under Temporal Correlations, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2018.

Soichiro Hidaka, Frédéric Jouault, Massimo Tisi. On Additivity in Transformation Languages. MODELS 2017, pp. 23-33, 2017.

Hiroyuki Kato, Yasunori Ishihara and Torsten Grust. DDO-Free XQuery. DBPL 2017, Article No. 4, 2017

7. ホームページ等

<http://www.biscuits.work/>