

2024 年度年次報告書

基礎理論とシステム基盤技術の融合による Society 5.0 のための基盤ソフトウェアの創出

2022 年度採択研究代表者

天笠 俊之

筑波大学 計算科学研究センター
教授

検証可能なデータエコシステム

主たる共同研究者:

石川 佳治 (名古屋大学 大学院情報学研究科 教授)

小口 正人 (お茶の水女子大学 基幹研究院 教授)

鬼塚 真 (大阪大学 大学院情報科学研究科 教授)

宮崎 純 (東京科学大学 情報理工学院 教授)

森嶋 厚行 (筑波大学 図書館情報メディア系 教授)

研究成果の概要

前年度の成果を受けて、検証可能なデータエコシステムを実現するための基礎理論および要素技術の研究開発を行なった。具体的には、以下の項目について研究を実施した。

1. メタデータマネージャの要素技術として、1) Apache Spark を用いた並列分散プロパティグラフ処理手法の開発、2) 検証可能な分散プロパティグラフストアに関する研究開発を行うとともに、スマート農業に関する実証についてデータセットの検討を進めた。
2. 高性能データリネージュのためのプロパティグラフに基づくスキーマについて検討を行った。訓練と推論の二面性を持つ機械学習モデルを容易に扱えるよう、プロパティグラフを利用したデータリネージュのスキーマを提案した。さらに大規模データリネージュへのデータ操作を効率的に行うための GPU を利用した高速化の研究も併せて実施した。
3. 企業間データ連携を前提とする検証可能なデータ共有プラットフォームとして、ブロックチェーンベースのモデルを検討し、Hyperledger Iroha を具体的な実装ターゲットとして評価を行った。
4. 品質を保証する近似的データベース問合せ処理、データストリーム処理における品質を考慮した近似処理に関する研究、近似的な最近傍問合せ処理の高速化、データベースにおける因果推論問合せ手法の開発に関する研究開発を行なった。
5. データベース修復・シミュレーションコア技術として、アプリケーションロジックとデータベース挙動との橋渡しを実現した統合的な What-if 分析フレームワーク Ultraverse、大規模言語モデル (LLM) を用いた個人移動生成のための新たな移動軌跡データシミュレーション手法、LLM を用いたデータ前処理 (エラー修復, 欠損値補完, データ統合) の高性能化、プライバシー保護型データ生成と価格評価等について研究開発を行なった。
6. ヒューマンインザループによる開世界データ信頼性向上技術として、ヒューマン・イン・ザ・ループ型のデータコンテンツと構造の有益性を検証するフレームワーク、ドメイン公理に基づくエンティティマッチングに関するアクティブラーニングのためのサンプリング手法の実証実験、ヒューマンコンピューテーションのためのデータ検証支援等について研究開発を行なった。

【代表的な原著論文情報】

- 1) Savong Bou, Toshiyuki Amagasa, Hiroyuki Kitagawa, "O(1)-Time Complexity for Fixed Sliding-Window Aggregation Over Out-of-Order Data Streams," IEEE Trans. Knowl. Data Eng., Vol. 36, No. 11, pp. 6745-6757, 2024.
- 2) Shinya Miura, Qiong Chang, Jun Miyazaki: k-Way In-Place Merge by CPU-GPU Cooperative Processing. IEEE ASAP 2024: 152-160, 2024.7
- 3) Haruka Hori and Masato Oguchi: "A Study of Blockchain-Based Metadata Management and Its Use for Data Verification" In Proc. the 12th International Symposium on Computing and Networking (CANDAR2024), Okinawa, Japan, November 2024.
- 4) Tianjia Ni, Kento Sugiura, Yoshiharu Ishikawa, Kejing Lu. "Guaranteeing an Exact Error Bound for Bounded Approximate Query Processing", Journal of Information Processing, Vol. 32, pp. 903-915, 2024.
- 5) Yunqing Ge, Jianbin Qin, Shuyuan Zheng, Yongrui Zhong, Bo Tang, Yu-Xuan Qiu, Rui Mao, Ye

Yuan, Makoto Onizuka, and Chuan Xiao. Privacy-Enhanced Database Synthesis for Benchmark Publishing. PVLDB, 18(2): 413 - 425, 2024.

- 6) Hiroyoshi Ito, Takahiro Koizumi, Ryuji Yoshimoto, Yukihiro Fukushima, Takashi Harada, Atsuyuki Morishima, "Inconsistency-driven Approach for Human-in-the-loop Entity Matching", Information Research, vol. 30, no. iConf(2025), March 2025