

2023 年度年次報告書  
AIP 加速課題  
2023 年度採択研究代表者

松井 勇佑

東京大学 大学院情報理工学系研究科  
講師

超高速データサイエンス基盤

主たる共同研究者:

天方 大地 (大阪大学 大学院情報科学研究科 助教)

塩川 浩昭 (筑波大学 計算科学研究センター 准教授)

西村 真衣 (オムロンサイニクエックス(株) リサーチアドミニストレイティブディビジョン シニアリ  
サーチャー)

## 研究成果の概要

本研究課題は様々なモーダルの高速探索技術を追求し、多モーダルにまたがる新しい技術を生み出すことを目指す。そして、実際に人々が使う基盤的な技術を創出することを目標とする。2023年度では、まず各モーダルにおける探索技術を追求し、IJCAI, ACMMM, ACL, NeurIPSなどのトップ会議への採択を果たした。そして、その社会への還元として近傍探索コンペへの参加および上位受賞、トップ会議での国際チュートリアル実施を果たした。そのうち代表的な3つを下記に述べる。

研究成果 1)にて最大密部分グラフを高速に検出する手法 **BnM** を開発した。本成果では実世界のグラフが持つ頻出構造を捉えることで、数千万件規模のデータにおいて従来法よりも2,300倍程度高速に計算可能にした。本成果はソーシャルデータ分析やAI創薬といったグラフ構造に紐づいた分析が重要となる広い研究分野で重要な要素技術であることから、人工知能分野のトップ会議のひとつであるIJCAI2023に採録された。

研究成果 2)にて、学習型ブルームフィルタの高速構築方式を提案した。ブルームフィルタとはデータの集合を省メモリに表現するデータ構造である。ブルームフィルタは基盤的なデータ構造であり、世界中で広く用いられている。本成果では、そのような基盤的なデータ構造の最新版である学習型ブルームフィルタについて、構築速度を700倍程度高速化した。この内容は機械学習分野におけるトップ会議であるNeurIPS2023に採録された。

研究成果 3)に示す通り、近似最近傍探索の国際コンペティション SISAP Indexing チャレンジにおける中規模データタスクに参加し、世界2位の成果を得た。本コンペにて提案した内容は現在の近傍探索におけるデファクト方式であるグラフ探索方式について、極めて実用的なパラメータチューニング方式を提案したものである。本研究課題におけるテーマである「実用的な方式」を実現し、それをコンペという形で示すことが出来た。

### 【代表的な原著論文情報】

- 1) Shohei Matsugu, Yasuhiro Fujiwara, Hiroaki Shiokawa, “Uncovering the Largest Community in Social Networks at Scale,” In Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI 2023), pp.2251-2260, Macao, China, August 2023.
- 2) Atsuki Sato and Yusuke Matsui, “Fast Partitioned Learned Bloom Filter”, NeurIPS 2023.
- 3) Yutaro Oguri and Yusuke Matsui, “General and Practical Tuning Method for Off-the-Shelf Graph-Based Index”, SISAP 2023 Indexing Challenge Report by Team Utokyo, 2023.