

高速かつ正確に関係性を理解するための グラフ分析技術

情報科学研究科 マルチメディア工学専攻

助教 佐々木 勇和

Researchmap <https://researchmap.jp/sasakiyuya>



研究の概要

グラフはモノの状態とそれらの関係性を表現するデータ構造です（棒グラフや円グラフのような図ではありません）。人間関係や、購買履歴、分子、道路網、知識（図1）などがグラフで表現可能であり、様々な分野で活用されています。例えば、新たな触媒探索のような物質開発やウェブショッピングにおける商品推薦などが実用的に使われています。私は大規模かつ多様なグラフデータに対して、(1) 効率的な管理と高速な検索を可能にするデータベース技術、(2) 新たな知見を発見するデータマイニング技術、(3) 正確な予測を行う深層学習技術を研究しています（図2）。

は様々な関係性が密接しており、関係性のグラフデータ化が進んでいます。これらのデータ化された関係性からモノの状態を予測すること、モノ同士の状態から関係性を予測すること、およびそれらを管理し、そこから新たな知見を発見することは世の中の理解と発展に必要です。大規模グラフの分析と管理を簡単に使えるプラットフォームを開発し、過去 / 現在 / 未来における大規模かつ多種多様な関係性を高速かつ高精度に分析できる技術の開発を進め、より快適かつ幸せな世界の実現に貢献していきます。

研究の背景と結果

我々の生活にはモノ単体の情報だけではなく、モノとモノの関係性が重要です。関係性を表すのに適したデータ構造がグラフであり、モノを節点、モノとモノの間の関係性を枝として表現します。グラフは様々な身近な応用に利用されています。例えば、ある人がある町の出身であるといった関係性や、ある国はある地域に属しているといった関係性、ある二つの商品は類似しているといった関係性を知識グラフとして構造化することで、Web 検索（例えば、Google 検索においてある人を検索した時に右側の表）や推薦システム（例えば、Amazon の関連商品推薦）に利用されています。他にも、分子データは原子を節点とし原子間の結合を関係性として表現することで、同じデータ構造をもつ分子の検索などに活用できます。ソーシャルネットワークサービスや道路網なども、我々の生活に密接しており、グラフの大規模化と多様化が進んでいます。

私はこのようなグラフデータに対して、管理・検索・発見・予測という4つの観点から研究開発を進めており、具体的に (1) 効率的な管理と高速な検索を可能にするデータベース技術、(2) 新たな知見を発見するデータマイニング技術、(3) 正確な予測を行う深層学習技術を研究しています（図2）。

研究の成果として、(1) 効率的な管理と高速な検索を可能にするデータベース技術においては、グラフデータの構造の特徴を事前計算することで、グラフデータベースにおける問合せの1000倍以上の高速化を達成しています。(2) 新たな知見を発見するデータマイニング技術においては、グラフデータ内の関係性の強さや特徴的な関係性を抽出することで、例外的な関係性や、さらにグラフデータ内の差別的なバイアスが含まれているかを評価することができる技術を開発しています。(3) 正確な予測を行う深層学習技術においては、大規模なグラフデータに適用可能な深層学習技術を開発し、精度を維持しながら、メモリ使用量が小さいかつ高速な学習を可能にしました。さらに、基礎開発だけではなく、これらのグラフの管理と分析技術を物質探索や医療AIなどの応用分野への適用も行い、他分野の研究者との共同研究も進めています。

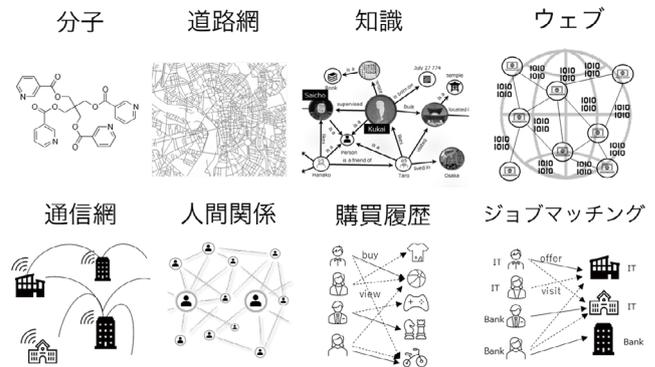


図1 グラフデータの応用例

コミュニティ検出 & リンク予測

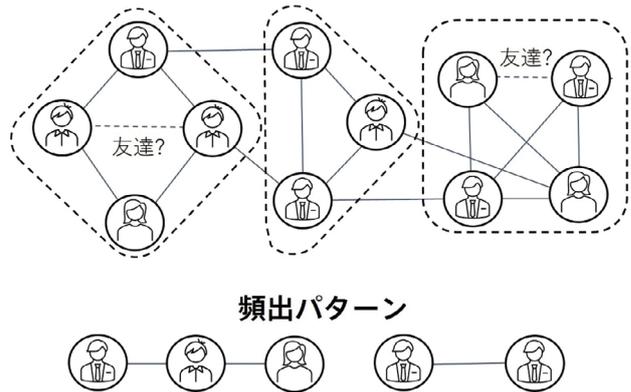


図2 グラフ分析の例

研究の意義と将来展望

ビッグデータ時代と呼ばれるように大規模かつ多様なデータを効率的に取り扱う技術の重要性が益々増えています。人間関係や、人と職、人が所有する物、Web のアクセス、国家間の協定など、我々の生活に

特許	
論文	Maekawa, Seiji; Noda, Koki; Sasaki, Yuya et al. Beyond Real-world Benchmark Datasets: An Empirical Study of Node Classification with GNNs. Conference on Neural Information Processing Systems (NeurIPS), Nov. 2022. doi: 10.48550/arXiv.2206.09144 Sasaki Yuya. Cost-constrained Minimal Steiner Tree Enumeration. Proceedings of International Conference on Information and Knowledge Management (CIKM), pp. 4439-4443, Oct. 2022. doi: 10.1145/3511808.3557570 Sasaki Yuya, George Fletcher, and Onizuka Makoto. Language-aware Indexing for Conjunctive Path Queries. Proceedings of International Conference on Data Engineering (ICDE), pp. 661-673, May 2022. doi: 10.1109/ICDE53745.2022.00054
参考URL	
キーワード	グラフニューラルネットワーク、ビッグデータ、データマイニング、データベース