

並列グラフマイニングアルゴリズムの開発

Development of Parallel Graph Mining Algorithms

塩川 浩昭

筑波大学 計算科学研究センター

1. 研究目的

我々のプロジェクトでは実世界の極めて大規模なグラフデータを対象に、並列計算環境を利用したマイニングアルゴリズムの開発している。本プロジェクト期間では並列グラフマイニングアルゴリズムの性能評価と計算コードの開発を目的とする。具体的には、昨年度開発した手法を対象として、(1) Oakforest-PACS 上でのプロダクトランおよび(2) Cygnus 上でのコード開発を行う。

2. 研究成果の内容

本年度の主要な研究成果は、(1)大規模なグラフデータを対象とした分散並列クラスタリング手法の開発ならびに (2)大規模なデータを対象としたコミュニティ検索手法の高性能化ならびに性能検証の 2 点である。以下のそれぞれの概要を述べる。

(1) 大規模グラフデータを対象とした分散並列クラスタリング手法の開発・性能検証

多種類のノードを持つヘテロジニアスグラフを対象としたクラスタリング手法 RankClus の高速化を行った。RankClus はランキングとクラスタリングを繰り返し行い、相互に補完することで従来の手法より正確で効率的な分析を可能にした手法である。しかしながら、RankClus はグラフに含まれるクラスタの数だけ部分グラフを作成し、全てに対してランキングとクラスタリングを実行する必要があり計算時間が膨大になってしまうという問題点がある。本研究期間では、RankClus の差分更新法を提案するとともに並列化を行うことで、精度を落とさず性能の大幅向上を達成した。

(2) 大規模データにおけるコミュニティ検索手法の高性能化ならびに性能検証

コミュニティ検索は問合せに対して特定の評価関数の値を最大化するコミュニティを見つけ出すグラフ分析手法であるが、従来提案された検索手法は大規模なデータに対して大きな実行時間を要する。本研究期間ではビームサーチにおける近傍解の高速列挙法を提案し、先行研究と同程度の精度を示しつつ、最大約 500 倍高速であることを確認した。

3. 学際共同利用が果たした役割と意義

近年、計算情報学の分野では利用可能なデータの規模が増加の一途をたどっており、学際共同利用を通じて提供される高性能な計算環境無くしては処理できない状況となっている。したがって、学際共同利用として実施した意義が大きい。

4. 今後の展望

今後はより大規模かつ多様な種類のデータに対して、本年度開発した成果の性能検証を深める。また、本研究を通じて獲得した並列化や高速化の手法を他のアルゴリズムへと適用し幅広い分析の高速化に発展させる予定である。

5. 成果発表

(1) 学術論文

- Kotaro Yamazaki, Shohei Matsugu, Hiroaki Shiokawa, Hiroyuki Kitagawa, "Fast and Parallel RankClus Algorithm based on Dynamic Rank Score Tracking," Journal of Information Processing , Vol.28, 12 pages, July 2020. (査読あり)
- Hiroaki Shiokawa, Yasunori Futamura, "Efficient Vector Partitioning Algorithm for Modularity-based Graph Clustering," Journal of Data Intelligence, Vol.1, No.2, pp.101-123, June 2019. (査読あり)
- Kotaro Yamazaki, Tomoki Sato, Hiroaki Shiokawa, Hiroyuki Kitagawa, "Fast and Parallel Ranking-based Clustering for Heterogeneous Graphs," Journal of Data Intelligence, Vol.1, No.2, pp.137-158, June 2019. (査読あり)
- Hiroaki Shiokawa, Tomohiro Matsushita, Hiroyuki Kitagawa, "Fast Affinity Propagation by Cell-based Indexing," Journal of Data Intelligence, Vol.1, No.1, pp.55-74, April 2019. (査読あり)
- Kotaro Yamazaki, Shohei Matsugu, Hiroaki Shiokawa, Hiroyuki Kitagawa, "Fast RankClus Algorithm via Dynamic Rank Score Tracking on Bi-type Information Networks," In Proceedings of the 21st International Conference on Information Integration and Web-based Applications and Services (iiWAS2019), pp.110-117, Munich, Germany, December 2019. (査読あり)
- Shohei Matsugu, Hiroaki Shiokawa, Hiroyuki Kitagawa, "Flexible Community Search Algorithm on Attributed Graphs," In Proceedings of the 21st International Conference on Information Integration and Web-based Applications and Services (iiWAS2019), pp.103-109, Munich, Germany, December 2019. (査読あり)

(2) 学会発表

該当なし

(3) その他

該当なし

使用計算機	使用計算機 に○	配分リソース※	
		当初配分	追加配分
Cygnus	○	5000	0
Oakforest-PACS	○	4800	0
※配分リソースについてはノード時間積をご記入ください。			