

発表概要

分散メモリ環境における並列グラフマイニングの実現に向けて

奥野 伸吾^{1,a)} 平石 拓^{2,b)} 中島 浩^{2,c)} 八杉 昌宏^{3,d)} 瀬々 潤^{4,e)}

2016年1月13日発表

本発表では、各頂点がアイテムの集合を持つようなグラフからユーザが指定した閾値以上の数の共通アイテム集合を持つ連結部分グラフを全抽出するグラフマイニングの、タスク並列言語 Tascell を用いた分散メモリ環境向けの実装について報告する。我々はすでに、この問題を解く探索アルゴリズム COPINE の共有メモリ環境向けの並列実装を提案している。この実装では、探索中に蓄えた知識に基づいて不要な探索を回避するため、1つの情報管理テーブルを全ワーカで共有する。しかし分散メモリ環境でこの手法を用いると、通信コストが大きくなってしまふ。そこで、ノードをまたぐワークスティールのタイミングで、そのノード間のテーブル情報の差分を送信する手法を実装した。また、COPINE ではタスク生成のコストが大きいため、ノード間のスティール回数の最小化を指向する従来の Tascell のスティール戦略では、ノード内で小さい仕事を多く盗み合うことによる性能劣化が大きくなってしまふ。そこで、ノード外へのタスク要求を増やすことにより大きな仕事を得やすくするスティール戦略の検討、実装も行った。これらの工夫により、一定の性能改善効果が得られることを確認した。

Towards Parallel Graph Mining in Distributed Memory Environments

SHINGO OKUNO^{1,a)} TASUKU HIRAISHI^{2,b)} HIROSHI NAKASHIMA^{2,c)}
MASAHIRO YASUGI^{3,d)} JUN SESE^{4,e)}

Presented: January 13, 2016

This paper reports a parallelized implementation of graph mining that extracts all connected subgraphs with common itemsets whose size is not less than a given threshold from a graph and itemsets associated with vertices of the graph, in distributed memory environments using the task-parallel language Tascell. For this problem, we have already proposed a backtrack search algorithm called COPINE and its parallel implementation in shared memory environments. In this implementation, all workers share a single table that contains the knowledge acquired during search in order to avoid useless search. This sharing method will crucially increase the cost for internode communications in distributed memory environments. Therefore, we implemented a sharing method where the difference of table information between computing nodes is sent when an internode work-steal occurs. There is another problem that the task creation cost is high for COPINE and thus the conventional work-stealing strategy in Tascell, which aims to minimize the number of internode work-steals, significantly degrades the performance since it increases the number of intranode work-steals for small tasks. Therefore, we implemented and evaluated work-stealing strategies that promotes workers to request tasks to external nodes to increase the opportunity to obtain large tasks. We confirmed that we can get the performance improvement using these methods.

¹ 京都大学大学院情報学研究科
Graduate School of Informatics, Kyoto University, Kyoto
606-8501, Japan

² 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto
University, Kyoto 606-8501, Japan

³ 九州工業大学大学院情報工学研究科
Department of Artificial Intelligence, Kyushu Institute of
Technology, Iizuka, Fukuoka 820-8502, Japan

⁴ 産業技術総合研究所
National Institute of Advanced Industrial Science and Tech-
nology, Koto, Tokyo 135-0064, Japan

a) shingo@sys.i.kyoto-u.ac.jp

b) tasuku@media.kyoto-u.ac.jp

c) h.nakashima@media.kyoto-u.ac.jp

d) yasugi@ai.kyutech.ac.jp

e) sese.jun@aist.go.jp